

# JOURNAL *of* EDUCATIONAL RESEARCH

## EDITORIAL BOARD

A. S. BARR, Chairman, Professor of Education, University of Wisconsin, Madison 6, Wisconsin

E. J. ASHBAUGH, Dean Emeritus, College of Education, Miami University, Oxford, Ohio. Teacher Education.

WILLIAM A. BROWNELL, Dean, College of Education, University of California, Berkeley, California. Educational Psychology, Learning and Teaching.

LEO J. BRUECKNER, Professor of Education, University of Minnesota, Minneapolis, Minnesota. Elementary Education.

ROBERT A. DAVIS, Head, Department of Psychology, George Peabody College for Teachers, Nashville, Tennessee. Evaluative Techniques.

HARL R. DOUGLASS, Director, School of Education, University of Colorado

Boulder, Colorado. Secondary Education.

A. K. LOOMIS, Director, School of Education, University of Denver, Denver, Colorado. Curriculum Construction.

M. J. Nelson, Dean of the Faculty, Iowa State Teachers College, Cedar Falls, Iowa. Teacher Education

WILLARD C. OLSON, Director of Research in Child Development, University of Michigan, Ann Arbor, Michigan. Child Development.

PERCIVAL M. SYMONDS, Professor of Education, Teachers College, Columbia University, New York City. Mental Hygiene, Guidance, and Pupil Adjustment.

---

VOLUME XLVII

September, 1953 - May, 1954

Copyright 1953 by  
DEMBAR PUBLICATIONS, INC.  
Madison, Wisconsin



## INDEX

VOLUME XLVII, SEPTEMBER, 1953 - MAY, 1954

### Listed Under Articles and Authors

- Achievement and Social Adjustment of Pupils in Combination Classes Enrolling Pupils of More Than One Grade Level, Joseph J. Adams, Oct 53, 151-156.
- Achievement, Aptitude, and Background of Liberal Arts and Science Students (General Curriculum) Deficient in High School Mathematics, Clarence Phillips, Nov 53, 169-180.
- A Comparative Study of 1939 and 1950 Achievement Test Results in the Hawthorne Elementary School in Ottawa, Kansas, Norris G. Blake, et al. Sept 53, 19-34.
- A Comparative Study of the Relationship Between the Quality of the Child's Language Usage and the Quality and Types of Language Used in the Home, Doris Ione Noel, Nov 53, 161-168.
- A Comparison of the Extent of Retention Accompanying the Use of Three Teaching-Testing Procedures, Lloyd L. Pathael, et al., Sept 53, 65-70.
- A Criticism of Studies Utilizing the "Index of Industriousness" as a Measure of Work Habits, Arvo E. Juola, Dec 53, 307-312.
- Action Research in Education, Stephen M. Corey, Jan 54, 375-380.
- Adams, Joseph J. Achievement and Social Adjustment of Pupils in Combination Classes Enrolling Pupils of More Than One Grade Level, Oct 53, 151-156.
- A Descriptive Analysis of 248 Non-High School Graduates Admitted to the University of Utah, Dean C. Andrew, April 54, 589-598.
- A Longitudinal Comparison of Freshman-Senior Standing, Leslie C. Carlin, Dec 53, 285-290.
- Amount of Free Reading Engaged in by Intermediate Grade Pupils Who have Viewed Television for One Year or More, Charles A. Wells et al., Feb 54, 473-478.
- Andrew, Dean C. A Descriptive Analysis of 248 Non-High School Graduates Admitted to the University of Utah, April 54, 589-598.
- Andrew, Gwen, et al., Teachers' Evaluations of the Mental Health of Their Pupils, April 54, 631-636.
- An Experiment in the Continuity-Type Question, R. R. Ashburn, et al., Nov 53, 201-210.
- An Investigation of Certain Factors Relating to the Nature of Children with Functional Defects of Articulation, Oliver W. Nelson, Nov 53, 211-216.
- A Preliminary Study of the Group Attitudes of Junior and Senior High School Students Towards Mathematics, William H. Malone, et al., April 54, 599-608.
- A Proposed Technique for the Analysis of Drop-Outs at a State College, Granville B. Johnson, Jan 54, 381-386.
- A Resume of Current Teacher Personnel Research, Glen Fulkerson, May 54, 669-682.
- Art in Writing for Educational Periodicals: The Main Body, J. R. Shannon, Mar 54, 489-504.
- Ashburn, R. R. et al., An Experiment in the Continuity-Type Question, Nov 53, 201-210.
- A Study of the Anti-Democratic Potential of Teachers, Stewart Jones et al., Sept 53, 1-18.
- A Study of Intercultural Adjustment Problems of Missionary Children, Howard S. Bretsch, April 54, 609-616.
- A Study of Peer Group Social Acceptability at the Fifth Grade Level in a Public School, Elva E. Kinney, Sept 53, 57-64.
- A Study of the Relationships Among Selected Language Abilities, Vergil H. Hughes, Oct 53, 97-106.
- Auston, John T. Rationale for Reorienting Educational Research, Jan 54, 337-350.
- A Technique for Evaluating Student Comments About Grading Practices, Lawrence P. Blum et al., Jan 54, 397-398.
- A Technique for a Statistical Interpretation of an Interest Survey Questionnaire, Sam S. Blanc, Nov 53, 223-230.
- Beach, Norton L. Research Goes Into Action, Jan 54, 351-360.
- Beckman, Milton W. Is General Mathematics or Algebra Providing Greater Opportunity to Attain the Recommended Mathematical Competencies? Mar 54, 505-512.
- Bedoian; Vaghars H. Social Acceptability and Social Rejection of the Underage, At-Age, and Overage Pupils in the Sixth Grade, Mar 54, 513-520.
- Blake, Norris G. et al., A Comparative Study of 1939 and 1950 Achievement Test Results in the Hawthorne Elementary School, Ottawa, Kansas, Sept 53, 15, UPON OR
- Blanc, Sam S. A Technique for a Statistical Interpretation of an Interest Survey Questionnaire, Nov 53, 223-230.
- Blum, Lawrence P. et al., A Technique for Evaluating Student Comments About Grading Practices, Jan 54, 397-398.
- Braden, Myron M. Former Students Evaluate Guidance, Oct 53, 127-134.
- Bretsch, Howard S. A Study of Intercultural Adjustment Problems of Missionary Children, April 54, 609-616.
- Bock, Carolyn E. Do We Need a New Basic Vocabulary for High School Latin, Dec 53, 301-306.
- Carlile, A. B. Predicting Performance in the Teaching Profession, May 54, 641-668.

- Carlin, Leslie C. A Longitudinal Comparison of Freshman-Senior Standing, Dec 53, 285-290.
- Carter, Robert S. Non-Intellectual Variables Involved in Teachers' Marks, Oct 53, 81-96.
- Charter, W. W. Research on School Board Personnel: Critique and Prospects, Jan 54, 321-336.
- Classroom Choice Status Related to Scores on Components of the California Test of Personality, Onas C. Scandrette, Dec 53, 291-296.
- Coladarci, A. P. et al., Research Priorities in Educational Administration, April 54, 625-630.
- Corey, Stephen M. Action Research in Education, Jan 54, 375-380.
- Correlations Meaning, Through Known Co-Variant Commonalities, Henry F. Dickenson, Sept 53, 71-76.
- Dickenson, Henry F. Correlations Meaning, Through Known Co-Variant Commonalities, Sept 53, 71-76.
- Dolch, E. W. et al., Vocabulary Tests and Depth of Meaning, Nov 53, 181-190.
- Do We Need a New Basic Vocabulary for High School Latin? Carolyn E. Bock, Dec 53, 301-306.
- Dreier, William H. Socio-Economic Status of Rural Youth Based on the Short Form of the Sewell Scale, April 54, 617-624.
- Droney, Margaret L. et al., Pupil Preference for Titles and Stories in Basal Readers for the Intermediate Grades, Dec 53, 271-278.
- Effective College Teaching: A Preliminary Causal Analysis, Robert M. Frumkin, et al., May 54, 683-694.
- Equivalence of the Pearson and Ayres Formulas of Correlation, George C. Kyte, Nov 53, 231-232.
- Former Students Evaluate Guidance, Myron M. Braden, Oct 53, 127-134.
- Fox, William H. et al., Validity of the Choice of Friends—Method of Measuring Social Adjustment, Jan 54, 389-396.
- Frumkin, Robert M. et al., Effective College Teaching: A Preliminary Causal Analysis, May 54, 683-694.
- Fulkeison, Glen, A Resume of Current Teacher Personnel Research, May 54, 669-682.
- Gray, W. S. Summary of Investigations: June 1 to June 30, 1953, Feb 54, 401-440.
- Green, Clinton W. Relationship Between Intelligence as Determined by Intelligence Tests and the Ability to Learn as Determined by Performance on Learning Tests, Nov 53, 191-200.
- Greenblatt, E. L. Two Additional Studies in the Dynamics of School Social Structure of Classroom Seating and School Dances, Dec 53, 261-270.
- Hughes, Vergil H. A Study of the Relationships Among Selected Language Abilities, Oct 53, 97-106.
- Hurlburt, Delpha, Relative Value of Recall and Recognition Techniques for Measuring Precise Knowledge of Word Meaning—Nouns, Verbs, Adjectives, April 54, 561-576.
- Improvement of Reading Skills Through Preparation of Materials, Rachel S. Sutton, Feb 54, 467-472.
- Influence of Word Analysis Upon Spelling Attainment, Joseph Jackson, Oct 53, 107-116.
- Instruction and Personality as Factors in Student Performance, Barry T. Jensen, Mar 54, 525-536.
- Is General Mathematics or Algebra Providing Greater Opportunity to Attain the Recommended Mathematical Competencies? Milton W. Beckman, Mar 54, 505-512.
- Item Analysis and Discriminative Value of Selected Wechsler-Bellevue Subtests, Edmund Med Dec 53, 241-260.
- Jackson, Joseph, Influence of Word Analysis Upon Spelling Attainment, Oct 53, 107-116.
- Jensen, Barry T. Instruction and Personality as Factors in Student Performance, Mar 54, 525-536.
- Johnson, Eric H. et al., Research Program of the Illinois Curriculum Program, Jan 54, 381-374.
- Johnson, Granville B. A Proposed Technique for the Analysis of Drop-Outs at a State College, Jan 54, 381-388.
- Jones, Stewart, et al. A Study of the Anti-Democratic Potential of Teachers, Sept 53, 1-18.
- Juola, Arvo E. A Criticism of Studies Utilizing the "Index of Industriousness" as a Measure of Work Habits, Dec 53, 307-312.
- Kelley, Dorothy Jones, Using Children's School Atypicalities to Indicate Ocular Defects, Feb 54, 455-466.
- Kinney, Elva E. A Study of Peer Group Social Acceptability at the Fifth Grade Level in a Public School, Sept 53, 57-64.
- Kyte, George C. Equivalence of the Pearson and Ayres Formulas of Correlation, Nov 53, 231-232.
- Logan, Virgil G. Mexico's Unified School System, Oct 53, 117-126.
- Ludlow, Herbert G. Trends and Issues in Standardized Testing, Dec 53, 279-284.
- Malone, William H. et al., A Preliminary Study of the Group Attitudes of Junior and Senior High School Students Towards Mathematics, April 54, 599-608.
- Mech, Edmund, Item Analysis and Discriminative Value of Selected Wechsler-Bellevue Subtests, Dec 53, 241-260.
- Mechanics of Questionnaire Construction, John E. Nixon, Mar 54, 481-488.
- Mees, John D. Teachers' Teaching Load in Illinois, Oct 53, 135-142.
- Mexico's Unified School System, Virgil G. Logan, Oct 53, 117-126.
- Nelson, Oliver W. An Investigation of Certain Factors Relating to the Nature of Children's Functional Defects of Articulation, Nov 53, 216.
- Nixon, John E. Mechanics of Questionnaire Construction, Mar 54, 481-488.
- Noel, Doris Ione, A Comparative Study of the Relationship Between the Quality of the Child's Language Usage and the Quality and Types of

- Language Used in the Home, Nov 53, 161-168.
- Notes on the Rating of Teacher Performance, David G. Ryans, May 54, 695-704.
- Non-Intellectual Variables Involved in Teachers' Marks, Robert S. Carter, Oct 23, 81-96.
- Oliver, W. A. Teachers' Education Beliefs Versus Their Classroom Practices, Sept 53, 47-56.
- Pathael, Lloyd L. et al., A Comparison of the Extent of Retention Accompanying the Use of Three Teaching-Testing Procedures, Sept 53, 65-70.
- Phillips, Clarence, Achievement, Aptitude, and Background of Liberal Arts and Science Students (General Curriculum) Deficient in High School Mathematics, Nov 53, 169-180.
- Phonic Knowledge and Its Relation to the Spelling and Reading Achievement of Fourth Grade Pupils, Mildred C. Templin, Feb 54, 441-454.
- Predicting Performance in the Teaching Profession, A. B. Carlile, May 54, 641-668.
- Prediction of Success in College Mathematics at Washburn University, William Seigel, April 54, 577-588.
- Primary Mental Abilities and Occupational Choices, Louis G. Schmidt, Dec 53, 297-300.
- Pupil Preference for Titles and Stories in Basal Readers for the Intermediate Grades, Margaret L. Droney, et al., Dec 53, 271-278.
- Rationale For Reorienting Educational Research, John T. Austin, Jan 54, 337-350.
- Relationship Between Intelligence as Determined by Intelligence Tests and the Ability to Learn as Determined by Performance on Learning Tests, Clinton W. Green, Nov 53, 191-200.
- Relationship of Primary Abilities Scores and Occupational Choices, Louis G. Schmidt et al., April 54, 637-640.
- Relative Value of Recall and Recognition Techniques for Measuring Precise Knowledge of Word Meaning—Nouns, Verbs, Adjectives, Delpha Hurlbut, April 54, 561-576.
- Research Goes Into Action, Norton L. Beach, Jan 54, 351-360.
- Research in Negro Colleges, J. Irving E. Scott, Mar 54, 521-528.
- Research on School Board Personnel: Critique and Prospects, W. W. Charters, Jan 54, 321-336.
- Research Priorities in Educational Administration, A. P. Coladucci et al., April 54, 625-630.
- Research Program of the Illinois Curriculum Program, Eric H. Johnson et al., Jan 54, 367-374.
- Ryans, David G. Notes on the Rating of Teacher Performance, May 54, 695-704.
- Scandrette, Onas C. Classroom Status Related to Scores on Components of the California Test of Personality, Dec 53, 291-296.
- Schmidt, Louis G. Primary Mental Abilities and Occupational Choices, Dec 53, 297-300.
- Schmidt, Louis G. et al., Relationship of Primary Abilities Scores and Occupational Choices, April 54, 637-640.
- Scott, J. Irving E. Research in Negro Colleges, Mar 54, 521-528.
- Seidman, Jerome M. et al., Teacher Likes and Dislikes of Student Behavior and Student Perceptions of These Attitudes, Oct 53, 143-150.
- Seigel, William, Prediction of Success in College Mathematics at Washburn University, April 54, 577-588.
- Shannon, J. R. Art in Writing for Educational Periodicals: The Main Body, Mar 54, 489-504.
- Social Acceptability and Social Rejection of the Underage, At-Age and Overage Pupils in the Sixth Grade, Vaghars H. Bedoian, Mar 54, 513-520.
- Socio-Economic Status of Rural Youth Based on the Short Form of the Sewell Scale, William H. Dreier, April 54, 617-624.
- Summary of Reading Investigations: July 1, 1952 to June 30, 1953, William S. Gray, Feb 54, 401-440.
- Sutton, Rachel S. Improvement of Reading Skills Through Preparation of Materials, Feb 54, 467-472.
- Teachers' Education Beliefs Versus Their Classroom Practices, W. A. Oliver, Sept 53, 47-56.
- Teachers' Evaluations of the Mental Health of Their Pupils, Gwen Andrew et al., April 54, 631-636.
- Teacher Likes and Dislikes of Student Behavior and Student Perceptions of These Attitudes, Jerome M. Seidman et al., Oct 53, 143-150.
- Teachers' Loading Load in Illinois, John D. Mees Oct 53, 135-142.
- Templin, Mildred C. Phonic Knowledge and Its Relation to the Spelling and Reading Achievement of Fourth Grade Pupils, Feb 54, 441-454.
- Trends and Issues in Standardized Testing, Herbert G. Ludlow, Dec 53, 279-284.
- Two Additional Studies in the Dynamics of School Social Structure of Classroom Seating and School Dances, E. L. Greenblatt, Dec 53, 261-270.
- Upon Use of Graphic Representation Problem-Solving at the College Level, H. Donald White, Sept 53, 35-46.
- Using Children's School Atypicalities to Indicate Ocular Defects, Dorothy Jones Kelley, Feb 54, 455-466.
- Validity of the Choice of Friends—Method of Measuring Social Adjustment, William H. Fox et al., Jan 54, 389-396.
- Vocabulary Tests and Depth of Meaning, E. W. Dolch et al., Nov 53, 181-190.
- Wells, Charles, et al., Amount of Free Reading Engaged in by Intermediate Grade Pupils

Who Have Viewed Television for One Year  
or More, Feb 54, 473-478.

White, Donald H. Use of Graphic Representation  
in Problem-Solving at the College Level, Sept  
53, 35-46.

#### RESEARCH NEWS

Nov 53: 239-240; Dec 53: 316-318; Jan 54: 399-  
400; Mar 54: 559-560; May 54: 713-715.

#### BOOK REVIEWS

Sept 53: 77-80; Oct 53: 157-160; Nov 53: 233-238;  
Dec 53: 313-315; May 54: 705-712.

#### ABSTRACTS

Dec 53: 319-320; Feb 54: 479-480; May 54: 716-  
720.

478273

# JOURNAL *of EDUCATIONAL RESEARCH*

---

Volume XLVII

September, 1953

Number 1

---

## A STUDY OF THE ANTI-DEMOCRATIC POTENTIAL OF TEACHERS

STEWART JONES  
EUGENE L. GAIER  
University of Illinois

### Introduction

THE PIONEER WORK of Lewin, Lippitt and White (4) demonstrating the impact of social climate on child behavior, has aided in modifying the earlier school emphasis on academic learning. At present, there is an increasing interest in psychological and social growth as a product of classroom interaction among teachers and students. This interest has gained research impetus by the reported investigations of the effect of teacher attitudes upon their pupils (3, 6, 7). Young (8) has asserted that many teachers are authoritarian in nature, have pre-determined attitudes, employ stereotyped and consequently mal-adaptive types of approaches to many problems, when he states that:

Another typical response is overemphasis upon obedience to authority. The teacher's own frustration at her inferior status, for example, may cause her unconsciously to find some compensation in strict rules and constant demands for obedience from her pupils. The prevalence of the parental pattern, anticipated by the community itself, supports the teacher in many of these personal forms of substitute satisfactions.

Closely related to this authoritarian pattern is the teacher who is insistent upon conformity to a wide range of external details or who is petty, resorts to continual nagging and whose discipline is constantly to the fore. Again, there is the teacher who is given over to repeated and almost omnipresent moralizing, not only in regard to the school subjects

but respecting the behavior and attitudes of the children before her.

The present study was initiated to determine the degree of authoritarianism in the beliefs and attitudes of public-school teachers. "Attitude" is here used in the manner defined by Kretch and Crutchfield (2),

As an enduring organization of motivational, emotional, perceptual, and cognitive processes with respect to some aspect of the individual's world.

We were here interested in investigating whether teachers differ in some of their basic attitudes from 1) those people preparing to teach (i. e., students working toward a degree in education), and 2) those college students not interested in the teaching profession. Specifically, the investigation was concerned with attempts to measure the degree of authoritarianism of the present subjects as could be inferred by the use of the measures of ethnocentrism and "anti-democratic potential" developed by the Berkeley group (1). We were also interested in the effects on teacher attitude of such factors as length of teaching experience, age, sex, and level of teaching.

### Method

The results of the present investigation are based on the use of the Ethnocentrism (E) and Pre-Fascistic (F) Scales, devised by the Berkeley Group (1), in Form 45 of their questionnaire<sup>1</sup>. It was believed that these instruments would lend themselves well to tapping the authoritarian character of our subjects. Levinson (1) has asserted that:

Ethnocentrism is based on a pervasive and rigid ingroup-outgroup distinction; it involves stereotyped negative imagery and hostile attitudes regarding outgroups, stereotyped positive imagery and submissive attitudes regarding ingroups, and a hierarchical, authoritarian view of group interaction in which ingroups are rightly dominant, outgroups subordinate. Sanford, et al., (1) describe the variables which go together to form the F Scale as a single syndrome: "a more or less enduring structure in the person that renders him receptive to antidemocratic propaganda." The F Scale is purported to measure such variables as:

1. This questionnaire was reprinted by permission from the Authoritarian Personality by T. W. Adorno, Else Frenkel-Brunswik, Daniel J. Levinson and R. Nevitt Sanford, Harper and Brothers, 1950.

TABLE I

CHARACTERISTICS OF THE THREE GROUPS OF THREE VARIABLES  
ASSUMED TO BE RELATED TO SOCIO-ECONOMIC STATUS

	Mean of Estimated Income in Ten Years <sup>1</sup>	Mean Number of Social Groups per Person <sup>2</sup>	Mean Number of Political and Professional Grps. with which each Member Affiliated	Mean in- come of Father <sup>3</sup>
Teachers (N = 57)	5270	.42	1.5	5200
Pre-Teachers (N = 76)	5000	.51	.21	5100
Non-Teachers (N = 62)	6600	.81	.26	7000

<sup>1</sup> Subjects asked to estimate their yearly income ten years hence. Figures shown are means of the amounts estimated.

<sup>2</sup> Subjects asked to list all organizations with which they were affiliated. These were arbitrarily subdivided into two groups, social and political-professional. The figures show the mean number of such groups listed by the respondents.

<sup>3</sup> These means are given to the nearest, even one-hundred dollars.

TABLE II

RELIGIOUS AFFILIATION AND RELIGIOUS ATTENDANCE OF THE THREE GROUPS  
(ALL FIGURES ARE PERCENTAGES OF TOTAL RESPONSES)

	Major Prot. Sect	Church Affiliation			Regu- larly	Church Attendance		
		Catholic	Jewish	Minor Prot. Sect		Often	Seldom	Never
Teachers (N = 57)	65	9	4	18	4	41	36	19
Pre-Teachers (N = 76)	59	21	7	5	8	26	24	35
Non-Teachers (N = 62)	57	12	16	12	3	25	20	44

. . . conventionalism, authoritarian submission, authoritarian aggression, anti-introception (opposition to the tender-minded) superstition and stereotypy, power and 'toughness', projectivity (projection outwards of unconscious emotional impulses), sex (exaggerated concern with sexual 'goings on').

Subjects—This questionnaire was administered to three groups of students enrolled at the University of Illinois during the Spring semester, 1951. Our population was selected from three distinct groups, in an attempt to test our initial hypothesis.

Group I, The Teacher (T) Group, consisted of 57 teachers enrolled in both university residence and extra-mural courses in the College of Education. These 26 women and 31 men ranged in age from 21 to 62 years, with a median age of 27.6 years. Teaching experience varied from one-half year to 34 years, with about an equal division of experience in elementary and high school teaching.

Group II, the Pre-Teacher (PT) Group, consisted of 76 students who had done supervised practice teaching in public schools for one semester, but who had not engaged in teaching on their own. They were preparing for professional positions for the coming year. These 21 women and 55 men in this group ranged in age from 20 to 41 years, with a median age of 23.8 years.

Group III, the Non-Teacher (NT) Group, consisted of 62 students from two classes in advanced journalism. These 19 women and 43 men, with an age range from 19 to 38 years and a median age of 22.4 years, were juniors and seniors in the College of Liberal Arts. These subjects differed from the Teacher Group in a number of seemingly important variables such as age, emphasis of instruction, etc. However the makeup of the Non-Teacher Group is comparable with that of the Pre-Teacher people, as the latter were seniors and first semester graduate students.

Questionnaires were administered to each group during a regular class period under comparable testing conditions. Anonymity was maintained for all groups, and their scores were not reported to the subjects.

The questionnaire consists of three sections: Section I contains background data such as age, sex, marital status, and religious and political affiliation. Section II, comprising the E and F Scales, consists of a series of statements, as:

- 1) Human nature being what it is, there will always be war and conflict.
- 2) Obedience and respect for authority are the most important virtues children should learn.
- 3) Zoot suiters prove that when people of their type have too much money and freedom, they just take advantage and cause trouble.

TABLE III  
PERCENTAGE OF PERSONS IN EACH GROUP AFFILIATED WITH EACH  
OF FOUR POLITICAL PARTIES

	Republicans	Democrats	Socialists	Progressives	Total per cent
Teachers (N = 57)	52	48	0	0	100
Pre-Teachers (N = 76)	51	46	0	3	100
Non-Teachers (N = 62)	70	28	2	0	100

TABLE IV  
RANGE OF SCORES ON THE E, F, AND COMBINED SCALES FOR  
THE THREE GROUPS

	<u>E Scale</u>			<u>F Scale</u>			<u>E and F Combined</u>		
	High Score	Low Score	Range	High Score	Low Score	Range	High Score	Low Score	Range
Teacher (N = 57)	106	28	78	153	55	98	154	89	154
Pre-Teacher (N = 76)	96	22	74	138	45	93	223	67	156
Non-Teacher (N = 62)	96	24	72	154	41	113	224	73	151

Twenty statements such as these compose the E Scale, and thirty similar items make up the F Scale. The subjects were instructed to mark statements with one of six values (+3 to -3) to indicate degree of agreement or disagreement with each item. The last section consists of a series of eight open-end questions, such as:

- 2) We all have impulses and desires which are at times hard to control, but which we try to keep in check. What desires do you often have difficulty in controlling?

Background factors (taken from Section I of the questionnaire) were coded numerically and included with the tabulation of each item of the E and F Scales. The responses to the open-end questions were classified under a number of qualitative categories derived by the authors and are treated in the next section (Part IID).

### Results

Since the sample used in this study was intact class groups, these results can be generalized only to the hypothetical population for which these groups may be considered random samples.

A. Background and Personal Data of Groups—The personal data collected in section I of the questionnaire showed certain marked differences among the three groups. The summary of this data is shown in Tables I, II, and III.

Although differences appear in religious affiliation among groups, these are not statistically significant. In the matter of attendance, however, differences between teachers and the other groups are significant. A chi square of the values under "Church Attendance" is 18.38, with six degrees of freedom, giving a probability of .01 that such differences could be accounted for through chance factors. This would indicate a trend for these teachers (as attendance is reported) to attend church more frequently than either of our other groups.

B. Comparison of Groups on Ethnocentrism and Anti-Democratic Potential—The range of scores for both scales was large in all three groups, as is shown in Table IV. The differences among the groups in mean scores is shown in Table V.

As is apparent in Table V, no statistically significant differences were obtained among these three groups on the E Scale. The within-group variability here exceeds the variability existing among the groups. The differences were more pronounced on the F Scale. Whereas the T and the NT groups had roughly comparable averages, the mean for the PT Group was considerably lower—presumably indicating a somewhat lower degree of anti-democratic potential than present in the other groups. The "t" value of the mean difference between the pre-teachers and non-teachers was 3.34, and between the pre-teachers

TABLE V  
DIFFERENCES AMONG THE THREE GROUPS ON THE E AND F SCALES

	Mean E Scale (Ethnocentrism)	S. D.	Mean F Scale Pre-Fascistic Tendency	S. D.	Mean Total E and F Scales	S. D.
Teachers (N = 57)	53.90	27.4	102.25	24.6	150.60	38.1
Pre-Teachers (N = 76)	50.50	19.3	93.20	25.2	139.15	32.2
Non-Teachers	51.04	17.2	106.02	20.4	149.99	41.0

TABLE VI

COMPARISON OF THE SCORES OF THE PRESENT STUDY WITH GROUP  
 SCORES REPORTED BY SANFORD'S ETAL ON  
 ON THE F SCALES, FORM 45

	N	Mean on F Scale
<u>Present Study</u>		
Teachers	57	3.36
Pre-Teachers	76	3.12
Non-Teachers	62	3.50
<u>Data of Sanford's etal</u>		
Testing Class Women	59	3.62
San Quentin Prisoners	110	4.73
Psychiatric Clinic Women	71	3.69
Psychiatric Clinic Men	50	3.82

and teachers, 2.11, giving probabilities of .01 and .04 respectively that these differences could occur by chance factors alone. When the scores of two scales were combined, the relative magnitude of the differences among the groups diminished.<sup>2</sup> However, the "t" value of the mean differences between the pre-teachers and teachers was 1.82 and between the pre-teachers and non-teachers, 1.69. Probabilities were thus slightly above the value necessary in rejecting the null hypothesis at the .05 level.

C. Comparison of These Groups with Berkeley Groups—In order to find where teachers stood with respect to larger, more varied groups than those included in this study, their scores were compared with those of the subjects tested by Sanford et al. (1). They reported the results of the use of Form 45 of the questionnaire with several groups, using item means rather than total means as presented here. Table VI contains the comparison of their findings with those of our groups.

D. Comparison of Teachers with Pre-Teachers on Specific Items—Since explanation based on scores often eclipses the points of difference on specific items, a comparison of the teachers with the pre-teachers on all the items of the F Scale was made. This comparison appeared justified since the mean differences were significant, and it indicated that approximately 40 per cent of the difference between total scores was attributable to only five of the 30 items included on the F Scale. These items, together with the mean differences and "t" values are contained in Table VII.

E. Comparison of Groups on Open-End Questions—As noted above, the responses to each of the eight open-end questions were grouped into a number of categories, e.g., on Item 1, "We all have times when we feel below par. What moods or feelings are the most unpleasant or disturbing to you?", responses were categorized as pertaining to:

- 1) worry about the future
- 2) concern over family matters
- 3) anxiety concerning realization of goals
- 4) depression about social injustice
- 5) depression because of other person's abuse of self
- 6) worry or guilt resulting from respondent's treatment of others
- 7) feelings of guilt unclassified
- 8) sex-related

In most categories, the three groups were quite similar in the frequency and kinds of responses. However, there were some apparent differences on each. With one or two exceptions, the teachers and

2. This is expected, since differences among groups on the E Scale were not pronounced.

pre-teachers were comparable in their response patterns.

Analysis of the open-end questions indicated the following differences among the groups:

1) As might be expected, pre-teachers and non-teachers are more concerned about the future (including military service) than are the teachers, due to age differences, past military experiences and dependents.

2) Apparently teachers are more apt than the other groups to worry over acts which have affected others (the fear that they have hurt others). Both teachers and pre-teachers appear to be more sensitive to both others' reactions toward them, and anxiety about their treatment of others than are the non-teachers.

3) Both teachers and pre-teachers reported instances of verbal aggression more often than non-teachers as either a "hard-to-control, impulse," or as a basis for feelings of depression.

F. A More Intensive Analysis of the Teacher Group—As shown in Table VI, none of the means of the present groups was as high on values of the F Scale as those groups reported by Sanford et al. The means were almost identical for the groups which were comparable (their testing class women, and our Non-Teacher Group). Moreover, means of the teachers' and non-teachers' scores closely approximated their results. But the variability within all groups was great. The range of the combined E and F scores for the T-group, 86 to 243, was very great.

In an attempt to note possible covarying functions of scores, a more thorough analysis was applied to the teachers. Our first question was whether magnitude of scores is dependent upon such factors as age, sex, and length of teaching. In general, this analysis produced no factors which could be identified as related to test scores. The obtained Pearson Product-Moment correlation coefficient between age and scores was .005. A correlation coefficient of .17 was obtained between length of teaching time and combined E and F scores. These were at a level expected by chance alone. When the group was divided according to sex (31 males and 26 females), the obtained means were 153.85 and 160.49 for women and men respectively. With standard deviations in the neighborhood of 40, these means are well within the range expected by chance.

When considerations of religion (as evidenced by reported church attendance) and political affiliation were inspected, differences were still within the range acceptable as due to chance variation. The correlation between the degree of religious attendance and total scores was .20. With 56 degrees of freedom, this value gives a "t" score of 1.78, which has a P value above .05, and is so low that conclusions or predictions about a relationship are unwarranted. The mean of that group of teachers (N=24) who professed affiliation with the Republican party was 166.66, and for the Democrats (N=22), 145.45.

TABLE VII

SCORES OF TEACHERS AND PRE-TEACHERS ON THE FIVE F-SCALE ITEMS FOR WHICH DIFFERENCES BETWEEN THE TWO GROUPS WAS GREATEST

Number	Item	Mean for Teachers	Sigma	Mean for Pre-Teach.	Sigma M. D.	t	P
2	Obedience and respect for authority are the most important virtues children should learn	3.82	1.92	3.26	2.01	.56	1.31
5	No weakness or difficulty can hold us back if we have enough will power	4.62	1.94	3.54	2.00	1.08	2.12 .01
13	A person who has bad manners, habits and breeding can hardly expect to get along with decent people	4.34	1.90	3.37	1.94	.97	2.26 .01
23	Young people sometimes get rebellious ideas, but as they grow up they ought to get over them and settle down	4.62	1.83	3.83	1.97	.79	1.88 .06
47	Homosexuals are hardly better than criminals and ought to be severely punished	2.62	1.62	2.04	1.59	.58	1.63 .08

The mean difference of 21.21 points was such that a "t" of 1.86 was obtained. This was not of sufficient magnitude to reject the null hypothesis.

Although teaching experience *per se* is not related to the magnitude of scores, it was found that higher scores were earned by those teachers who had taught in both elementary and high school. The total score mean for the group which had taught only at one school level ( $N=45$ ) was 147.11, while those who had taught in both ( $N=12$ ) averaged 179.33. The mean difference of 32.22 points resulted in a "t" value of 2.64, significant at about the .02 level with 27 d.f.

The top ten and the bottom ten scores (on combined E and F Scales) were selected, and a comparison made between the backgrounds of the high-score and the low-score groups, for were particularly interested in the characteristics of the extremes of the distribution of the Teacher Group. A comparison of these extremes is shown in Table VIII.

Teachers with high scores seemed to differ from the low-scoring group in political affiliation, and in their expression of religious affiliation. A Chi Square of the values shown under "Political Affiliation" is 5.00, which with one degree of freedom gives a P of .05 that this is a chance difference; and the Chi Square of 9.6 for the values under religious affiliation a P of less than .01 that such a chance difference. It may also be noted that high scorers had more elementary teaching experience, and had taught slightly longer than the low scorers.

### Discussion

If the obtained scores are a true representation of authoritarianism, the range of scores produced with the groups of the present study is so great that little difficulty will be met in ordering such persons on this dimension. However, several considerations preclude our acceptance of these scores at their face value. In the first place, some of the present subjects reported, after taking the test, that they were able to "see-through it." They could see that it was an attempt to tap prejudices and stereotyped thinking, and they could accordingly slant their responses at will. Secondly, there is probably some discrepancy between academic and working attitudes. It seemed apparent that students were responding as they thought they should respond, rather than as they actually felt. The teacher may know the kind of attitudes he as a teacher should have and slant his responses accordingly.

The statistically significant differences among our groups on the prefascistic scale is not explainable in terms of any of the collected background factors. Perhaps the PT Group, all full-time students

TABLE VIII

CHARACTERISTICS OF TEACHERS WITH THE TEN HIGHEST SCORES  
 ON THE E AND F SCALES COMBINED VERSUS THE  
 TEN WITH THE LOWEST SCORES

Group	Sex	Mean Age	Mean Annual Income	Mean of Est. Income in Ten Yrs.	Polit. Affil.		Relig. Affil. <sup>1</sup>		Mean Relig. Atten. <sup>2</sup>	Tot. Yrs. Element.	Tot. Yrs. High Sch.	Mean of Tot. Yrs.
					Dem.	Rep.	Prot.	Named Uncl. Sect				
10 High Scores	5-F				3	5	1	9	3.4	41.5	19	7.8
10 Low Scores	5-M	34.2	2577	5000	7	1	6	4	2.7	10.5	34	5.2

<sup>1</sup> Under "Named Sect" are figures for those persons who specified Methodist, Lutheran, etc. Those who simply wrote "Protestant" are listed as "Prot. Uncl."

<sup>2</sup> These means were obtained by converting to numerical scores the four possible categories of response: Regularly 4, Often 3, Seldom 2, Never 1.

in education, had given sufficient consideration to these important attitudes in their course work resulting, at least in written expression, that their academic attitudes were more liberal. Perhaps having become familiar with the origin and nature of such attitudes and scales for measuring them, they were less naive than the other groups. A look at the items on the F Scale which were higher for the T Group than the Pt Group seems to give a cluster of attitudes which have common elements. Two of the items, 2 and 23 (See Table V) deal with attitudes about the behavior of children; and two items, 13 and 47, deal with aspects of customs and mores. This may indicate that teaching experience, or selection has operated to make some teachers more inflexible in their perception and judgments about children, customs and mores. Further research may possibly identify the extent and source of these attitudes.

Teachers' relatively high scores may simply reflect that their attitudes have grown more similar to the attitudes of their communities, whereas the pre-teachers, as full time students in a university context still retain the more liberal attitudes of the college surroundings. As previously noted, even the Teacher Group is somewhat lower on the F Scale than somewhat comparable groups of the Berkeley studies. This is contrary to what might be expected, as we might well assume that being in a role of authority (the central figure in a classroom year after year—the person whose decisions in many cases are final judgments) would lead to more generalized authoritarian attitudes. We suggest that one of the following explanations may account for the comparatively low scores of the present group of teachers:

- 1) Certain authoritarian attitudes are specific to situations, i. e., there is a small relationship between classroom authoritarianism and general authoritarianism as measured by the E and F Scales.<sup>3</sup>
- 2) The classroom situation is not as authoritarian a setup as we have been led to believe—possibly due to the current emphasis in educational literature of the work of group dynamics and student-centered techniques.

- 3) Teachers' acquaintance with the kinds of attitudes they believe the investigators would want them to have may distort the present results.

Responses to open-end questions, in general, indicated differences among the groups which were consistent with what was already known about the differences in the groups' backgrounds. However, one added difference is the apparently greater sensitivity to the feelings of others which was evidenced in the responses of teachers and pre-teachers.

3. Research designed to test this hypothesis is under way by the present investigators.

The fact that various characteristics of teachers, such as age, length of teaching, level of teaching, religiosity, and political affiliation were seemingly not related to authoritarianism, as measured by the questionnaire, appears as evidence that such factors are but surface manifestations of a more basic attitude structure.<sup>4</sup> Indirect evidence for this is found in the fact that teachers who have taught at both levels of school have higher degrees of authoritarianism than those who have taught at one level. If it may be assumed that moving from one level to another is an indication of job dissatisfaction, we may assume that one of the important factors back of such attitudes might be more basic feelings about work and other areas. This suggests that the etiology of these authoritarian attitudes must be sought in more basic perceptions of the self, family, work and community.

### Summary and Conclusions

An Ethnocentric and Pre-Fascistic questionnaire (Form 45) devised by the Berkeley Group, for tapping authoritarian attitudes, was administered to three groups, teachers, pre-teachers, and non-teachers, all students at the University of Illinois. Data provided by the questionnaire included background factors, scores on an ethnocentric and pre-fascistic scale, and responses to a series of eight open-end questions. Added to the questionnaire was a section designed to reveal the amount and types of teaching experience. Findings were as follows:

1. Comparison of the three groups revealed differences in each of the above areas. In factors of background, it was found that both pre-teachers and teachers had a lower level of aspiration in terms of salary ten years hence than the non-teachers. Also, the non-teachers were, according to their reports of their fathers' incomes, from a higher economic level. There was a significant difference in the number and types of social and professional-political groups with which the subjects were affiliated. In the matter of religious affiliation, it was found that teachers gave a report of considerably more frequent church attendance than did the other groups. The non-teachers, reported 70 per cent affiliation with the Republican Party, while the pre-teachers and teachers reported about 50 per cent Republican and 50 per cent Democratic affiliation. Even though the groups were qualitatively quite different in a number of background factors, these observed differences did not manifest themselves as being related to individual tests scores.

4. Another explanation would be that the scales are not valid measures of authoritarianism. Methods of validation of the instrument have been sharply criticized by Luchins (5).

2. No significant differences were found among the groups on measured ethnocentrism. On anti-democratic potential, however, the pre-teachers were significantly lower than either of the other groups. Scores for all groups were slightly lower than those reported by the Berkeley group for somewhat comparable groups. In no case were differences great enough to warrant general conclusions about the attitude structure of the sample.

3. The items discriminating most between teachers and pre-teachers appeared to be concerned with social customs and mores, as well as with pre-determined evaluation of children's behavior. Comparison of scores with background factors of teachers provided no significant relationship between authoritarianism and age, length of teaching, school level taught, frequency of church attendance, or political affiliation. Those who had taught at both elementary and high-school levels made significantly higher scores than those who had taught at but one grade level.

4. In a comparison of the ten high-scoring teachers (high ethnocentrism and anti-democratic potential) with the ten lowest, the upper group was higher in religious attendance, amount of elementary school experience, and in total years taught. In addition, more of the high scorers professed affiliation with the Republican party.

5. In an overall appraisal of results, it must be noted that no marked differences between these groups in the degree of authoritarianism as measured was found. No conclusive evidence was found to show that teachers are significantly higher (or lower) than other groups in this dimension of personality.

The question as to the effect of such attitudes upon classroom behavior is a crucial one, in the face of present world conditions, and research to find relationships between these general attitudes and classroom attitudes is under way.

#### BIBLIOGRAPHY

1. Adorno, T. W.; Frenkel-Brunswik, Else; Levinson, Daniel J.; Sanford, R. Nevitt, The Authoritarian Personality, New York: Harper and Brothers, 1950.
2. Kretch, David and Crutchfield, R. S., Theory and Problems of Social Psychology, New York: McGraw-Hill, 1948.
3. Laycock, S. R., "Effect of Teacher's Personality on the Behavior of Pupils," Understanding the Child, 19, 50-55, 1950.

4. Lewin, Kurt, Lippit, R. and White, R. K., "Patterns of Aggressive Behavior in Experimentally Created Social Climates", Journal of Social Psychology, 10, 271-299, 1939.
5. Luchins, Abraham, S., "Personality and Prejudice: A Critique", Journal of Social Psychology, 32, 79-94, 1950.
6. Symonds, P. M., "Reflections on Observations of Teachers," Journal of Educational Research, 43, 1950, 688-696.
7. Witty, P. A., "Analysis of the Personality Traits of the Effective Teacher," Journal of Educational Research, 40, 1947, 662-671.
8. Young, Kimball, Personality and Problems of Adjustment, New York: F.S. Crofts and Co., 1946.

# A COMPARATIVE STUDY OF 1939 AND 1950 ACHIEVEMENT TEST RESULTS IN THE HAWTHORNE ELEMENTARY SCHOOL IN OTTAWA, KANSAS \*

NORRIS F. BURKE  
Hawthorne Elementary School  
Ottawa, Kansas

KENNETH E. ANDERSON  
University of Kansas  
Lawrence, Kansas

## The Problem

MANY PEOPLE say that the schools of today are not doing a good job of teaching the "fundamentals" or "three R's". They contend that children in the "good old days" really learned how to read, write, spell and cipher.

The questions that follow are: "Do children in elementary schools today master the fundamental subjects as well as their parents and grandparents did? Does modern education neglect the three R's?"

The authors have been concerned with these questions for sometime. This was an attempt to evaluate the results of the teaching done in one grade school. The purpose of the study was to make a comparison of pupils' achievement test scores in the Hawthorne Elementary School, Ottawa, Kansas, for the years 1939 and 1950.

## The School

The Hawthorne Elementary School has a kindergarten and the first six grades. It is located in a town of 10,000 population in an average community. The enrollment in 1939 was 324 children and in 1950 there were 307 children enrolled.

The curriculum of the school in 1939 could well be classed as the formal subject matter type where textbooks were used predominately.

\* The title of the Master's Thesis submitted by Mr. Burke in partial fulfillment of the requirements for the degree of Master of Science in Education, Kenneth E. Anderson, adviser.

The curriculum in 1950, while still using subjects and textbooks, was enriched by the use of educational films, filmstrips, radio programs, field trips and other visual education aids to teaching. In 1950, less stress was put on the memorization of facts and more time spent on the development of attitudes, work habits, and the social living of children.

### The Pupils

The pupils who attended the school were from homes of the laboring class of people as well as from the middle class of people.

The 1950 pupils had kindergarten experience, whereas, the majority of the 1939 group had not been to kindergarten.

The 1939 group was slightly older than the 1950 group in most of the grades; however, the difference in age between the two groups was not enough to be significant.

Table I shows the number of pupils in the study in each grade for the two years, 1939 and 1950. Records were not available for all of the pupils for both years of the study.

The intelligence of the pupils was an important factor in the study and the differences in intelligence between the groups were adjusted for in the calculations for most of the comparisons made. The same intelligence tests were given the 1950 groups as had been given in 1939 with the exception of grades two and three.

### The Teachers

Three of the teachers in 1939 did not have a college degree and the same was true in 1950. Seven of the teachers in 1950 would be called beginning teachers as they had two or less years of experience; whereas, all but one of the 1939 teachers had nine or more years of experience.

### The Achievement Test

This study was based on the results of the Metropolitan Achievement Tests given to the 162 pupils in grades one to six on May 12, 1939 and 216 pupils in these grades who took the same tests but different forms on May 16, 1950.

All scores were recorded as grade equivalents on the class record sheets for the pupils who took the tests.

The Metropolitan Achievement Tests given in 1939 were the old forms, whereas those given in 1950 were the new forms. The old forms were not available to give in 1950.

September, 1953)

BURKE-ANDERSON

21

TABLE I

**NUMBER OF STUDENTS IN THE VARIOUS GRADES  
IN THE STUDY**

Year	Grade						Total
	I	II	III	IV	V	VI	
1939	22	18	35	36	19	32	162
1950	27	30	37	42	35	45	216

TABLE II

**SUMS OF SQUARES FOR ANALYSIS OF VARIANCE  
AND COVARIANCE, GRADE FIVE, READING**

Year	Number	$\Sigma Y^2$	$\Sigma YZ$	$\Sigma Z^2$
1939	19	4771.7894	205.9421	14.3115
1950	35	13624.1714	546.6914	29.6754
		18395.9608	752.6335	43.9869
Total	54	18561.4814	772.1592	46.2903



In the Booklet of Norms are found tables of corresponding scores on old and new forms of the Metropolitan Achievement Tests. These have been provided in order to enable the test user to express results on the old Forms A, B, C, D, and E of these tests in terms of scores on the new Forms R, S, T, U, and V, and vice versa.<sup>1</sup>

All the grade equivalents of the 1950 group on the new Forms R-V of the achievement test were converted to equal grade equivalents on the old Forms A-E by means of these tables given in the Booklet of Norms.<sup>2</sup>

### The Statistical Analysis

The hypothesis to be tested was that there was no significant difference in mean achievement test results for students of the 1939 group in the subjects of their grade as compared with the mean achievement test results for the students of the 1950 group for the same subjects of the same grade.

The basic assumptions which had to be satisfied before the analysis of variance and covariance tool could be applied were.<sup>3</sup>

1. that the variables under consideration must be approximately normally distributed in the population.
2. that there was no difference between the standard deviations of the 1939 groups and the 1950 groups with respect to the dependent variable.
3. that there was no difference between the within partial regression coefficients of the 1939 groups and the 1950 groups.

In order to test the first basic assumption, sample distributions were made for arithmetic and intelligence in grade four. These two distributions were found to be almost normal in regard to skewness and kurtosis. The histograms were uni-modal and appeared to be bilaterally symmetrical. Hence, we can say that in the case of intelligence and arithmetic for grade four, the departure from normality

<sup>1</sup> Booklet of Norms. Metropolitan Achievement Tests, World Book Company, Yonkers-on-Hudson, New York, p. 5.

<sup>2</sup> Ibid., p. 7-14.

<sup>3</sup> Jim Schunert, "The Association of Mathematical Achievement with Certain Factors Resident in the Teacher, in the Teaching, in the Pupil, and in the School," Journal of Experimental Education, 19, 3: 222-224; March, 1951.

was not great. Fisher<sup>4</sup> has proved that for curves that exhibit only a moderate departure from normality, the efficiency remains reasonably high. For the t test and F test used later in the analysis, it has been shown that no serious error for a slight departure from normality is introduced in the significance levels.<sup>5</sup> Thus in grade four, the departures from normality were not too marked. It is reasonable to expect much the same kind of distributions in the case of the scores in the other grades.

The conclusions based upon the statistical techniques used must be tempered by the fact that normality of each of the distributions was not rigorously tested.

The second and third basic assumptions were tested by using the Welch-Nayer L<sub>1</sub> test.

Where all the basic assumptions were met, the technique of analysis of variance and covariance was applied holding intelligence constant. When the F values were significant, the means were adjusted for the differences in intelligence.

The analysis of variance and covariance technique could not be applied to all of the subjects of all of the grades as both assumptions were not met. When this happened the Cochran-Cox technique was used.

A t test was made when the test for homogeneity of variances was accepted but the test for homogeneity of regression was rejected.

The twenty-seven comparisons made in this study were carried on in the manner described above. Throughout the study the null hypothesis was in use. The use of the designatory word "accept" therefore was interpreted to mean that the test had in fact demonstrated no difference, while the use of the term "reject" indicated that there was a difference significant at the specified level of confidence.

#### Sample Calculations for Analysis of Variance and Covariance

A sample calculation for analysis of variance and covariance where all assumptions were met is given for grade five, reading, in Tables II, III and IV. Table V shows the adjusted means for grade one, reading.

<sup>4</sup> R. A. Fisher, "On the Mathematical Foundations of Theoretical Statistics," Philosophical Transactions of the Royal Society of London, A, 222 (1922), p. 309-368.

<sup>5</sup> W. G. Cochran, "Some Consequences When the Assumptions for the Analysis of Variance are not Satisfied," Biometrics, 360: 22-28; March, 1947.

TABLE III

ADJUSTMENT TABLE FOR ANALYSIS OF COVARIANCE,  
GRADE FIVE, READING

Year	Correction	$\Sigma Z^2$
1939	8.8881	5.4234
1950	21.9368	7.7386
		(13.1620)
	30.7924	13.1945
Total	32.1218	14.1685

TABLE IV

ANALYSIS OF VARIANCE COVARIANCE OF FINAL SCORE WITH  
INTELLIGENCE CONSTANT, GRADE FIVE, READING

D. F.	S. S.	M. S.	F.	Hypothesis
51	13.1945	.2587		
1	.9740	.9740	3.76	P < .05
52	14.1685			

TABLE V  
ADJUSTED MEANS FOR GRADE ONE, READING

Year	N	Mean I. Q.	Diff. From Grand Mean I. Q.	Mean on Reading	Correction	Adjusted Means
1939	22	104.36	+2.86	2.59	.04	2.63
1950	27	109.56	-2.34	1.96	-.03	1.93
Sum	49	107.22				

Results of the Study

The results of the study where the analysis of variance and covariance technique was used are given in Table VI and VII. Tables VIII and IX show the results found where the Cochran-Cox method was used.

An examination of Table VI reveals that none of the F values were found to be significant at either the five or one per cent levels. Thus, in these subjects in their respective grades where the technique of analysis of variance and covariance was applied to the grade equivalents for these tests used, we concluded that there was no significant difference in mean grade equivalents for the test results in: (1) arithmetic, spelling and reading for grade two, (2) reading for grade three, (3) English and spelling for grade four, (4) reading, English and spelling for grade five, and (5) history and geography for grade six.

An examination of Table VII reveals that the F values for these different subjects of these different grades were significant at either the five per cent level or the one per cent level as indicated. The mean grade equivalents were adjusted for the differences in intelligence, which was held constant by the technique of analysis of variance and covariance. The adjusted mean grade equivalents were higher for the 1939 group than for the 1950 group in the following subjects and grades: (1) numbers and reading, grade one, (2) arithmetic, grade three, (3) literature, grade five, and (4) arithmetic and literature, grade six. The adjusted mean grade equivalents were higher for the 1950 group than for the 1939 group in the following subject and grades: (1) reading, grade four and (2) history and geography, grade five.

Thus, on the basis of the sample used, holding intelligence constant, the conclusions were:

1. The 1939 group attained a higher level of achievement than the 1950 group did in numbers and reading, grade one; arithmetic, grade three; literature, grade five; and arithmetic and literature, grade six.
2. The 1950 group attained a higher level of achievement than the 1939 group did in reading, grade four, and history and geography, grade five.

Table VIII and IX show the results of the study where the comparisons were made by the Cochran-Cox method.

An examination of Table VIII reveals that in grade three and six, the observed t was less than the criterion t for intelligence. Thus, the null hypothesis was accepted. The groups were homogeneous with respect to intelligence. The observed t for spelling, grade three, fell between the five and one per cent levels of the criterion t. The ob-

TABLE VI

COMPARISON OF GRADE EQUIVALENTS WITH  
INTELLIGENCE QUOTIENTS HELD CONSTANT

Test	Grade	Group	D. F.	F. *
Arithmetic	II	1939	1-45	0.09
		1950		
Spelling	II	1939	1-45	0.03
		1950		
Reading	II	1939	1-45	2.51
		1950		
Reading	III	1939	1-69	0.19
		1950		
English	IV	1939	1-75	0.15
		1950		
Spelling	IV	1939	1-75	0.11
		1950		
Reading	V	1939	1-51	3.76
		1950		
English	V	1939	1-51	. .
		1950		
Spelling	V	1939	1-51	. .
		1950		
History	VI	1939	1-74	. . . 89
		1950		
Geography	VI	1939	1-74	3.71
		1950		

\*None of the F values for the subjects in this table were significant at either the five or one per cent levels.

TABLE VII

COMPARISON OF GRADE EQUIVALENTS WITH  
INTELLIGENCE QUOTIENTS HELD CONSTANT

Test	Grade	Group	D. F.	F.	Adjusted means if F was significant
Numbers	I	1939	1-46	4.58*	2.84
		1950			2.67
Reading	I	1939	1-46	51.93**	2.63
		1950			1.93
Arithmetic	III	1939	1-69	15.57**	4.05
		1950			3.60
Reading	IV	1939	1-75	36.90**	5.12
		1950			5.80
History	V	1939	1-51	6.54*	5.12
		1950			5.64
Geography	V	1939	1-51	21.47**	5.14
		1950			5.93
Literature	V	1939	1-51	54.36**	5.82
		1950			5.78
Arithmetic	VI	1939	1-74	18.81**	7.32
		1950			6.53
Literature	VI	1939	1-74	4.22*	7.14
		1950			6.83

\* Significant at five per cent level

\*\*Significant at one per cent level

served t for spelling, reading and English for grade six were all larger than the criterion t at the one per cent level. Thus, for spelling in grade three and for spelling, reading and English for grade six, the null hypothesis was rejected. The observed t for language, grade three, was less than the criterion t. Thus for language, grade three, the null hypothesis was accepted.

The conclusions for the Cochran-Cox method on the basis of the sample used were:

1. The 1939 and the 1950 group in grade three and in grade six were homogeneous with respect to intelligence.
2. In grade three, the 1939 group attained a higher level of achievement in spelling than did the 1950 group.
3. The mean grade equivalents for the two groups, 1939 and 1950, in grade three, language, did not differ significantly. The two groups had made equal achievement in language.
4. In grade six, the 1939 group attained a higher level of achievement than the 1950 group in spelling, reading, and English.

An examination of Table IX reveals that the t ratios calculated on the intelligence scores for grade four and five were not significant. The two groups, 1939 and 1950, for grade four and for grade five were homogeneous with respect to intelligence. In arithmetic, grade four, the observed t was greater than the criterion t so the null hypothesis was rejected. In arithmetic, grade five, the observed t was less than the criterion t so the null hypothesis was accepted.

The conclusions, based on the sample used, with the t test for intelligence quotients and the Cochran-Cox method of comparison for the subject fields, were:

1. The two groups, 1939 and 1950, for grade four and for grade five were homogeneous with respect to intelligence.
2. The 1939 group in grade four attained a higher level of achievement than the 1950 group in arithmetic. This was not true in grade six where the difference in mean grade equivalents was not significant.

### Conclusions

It is wise to determine the significance of the standard (B's) partial regression coefficients for the independent variables prior to the use of the analysis of variance and covariance technique. If the B or B's are not significant, the independent variable or variables are not held constant. This was not done with the independent variable, intelligence, in this study.

TABLE VIII

COMPARISON OF INTELLIGENCE QUOTIENTS AND GRADE  
EQUIVALENTS BY COCHRAN-COX METHOD

Test	Grade	Group	Number	Variance Ratio	Mean	Difference in Means	Observed t	Criterion t
Intelligence	III	1939	35	2. 02*	102. 54	4. 76	1. 65	2. 03 (. 05 level)
		1950	37		97. 78			
Spelling	III	1939	35	2. 61**	3. 75	. 36*	2. 06	2. 03 (. 05 level)
		1950	37		3. 39			2. 72 (. 01 level)
Language	III	1939	35	3. 96*	3. 78	. 22	1. 11	2. 03 (. 05 level)
		1950	37		4. 00			
Intelligence	VI	1939	32	4. 20**	114. 31	5. 05	1. 36	2. 01 (. 05 level)
		1950	45		109. 26			
Spelling	VI	1939	32	2. 72**	6. 74	. 76**	3. 30	2. 71 (. 01 level)
		1950	45		5. 98			
Reading	VI	1939	32	3. 46**	7. 24	. 56**	2. 96	2. 70 (. 01 level)
		1950	45		6. 68			
English	VI	1939	32	2. 83**	7. 95	. 98**	4. 83	2. 70 (. 01 level)
		1950	45		6. 97			

\* Significant at five per cent level  
\*\* Significant at one per cent level

Note: If observed t is less than criterion t,  
the hypothesis is accepted.

TABLE IX

COMPARISON OF INTELLIGENCE QUOTIENTS AND GRADE  
EQUIVALENTS BY COCHRAN-COX METHOD

Test	Grade	Group Number	Variance Ratio	Mean	Difference in Means	t Ratio	Observed t	Criterion t
Intelligence	VI	1939	36	1.02	105.44	4.94	1.74	
		1950	42		100.50			
Arithmetic	IV	1939	36	7.71**	5.10	.45**		4.09
		1950	42		4.65			2.72 (.01 level)
Intelligence	V	1939	19	1.51	98.10	3.67	.684	
		1950	35		101.77			
Arithmetic	V	1939	19	3.56**	5.33	.29		1.26
		1950	35		5.04			2.09 (.05 level)

\* Significant at five per cent level  
\*\*Significant at one per cent level

Note: If observed  $t$  is less than criterion  $t$ ,  
the hypothesis is accepted.

All conclusions based on t tests or Chochran-Cox tests must be tempered by the fact that intelligence was not held constant, even though the t tests or the Cochran-Cox tests showed the groups to be homogeneous with respect to intelligence.

Thirteen of the comparisons made showed there was no significant differences in mean grade equivalents for the two groups, 1939 and 1950. Eleven of the comparisons made showed significant differences in favor of the 1939 groups, while three comparisons showed significant differences in favor of the 1950 groups.

The purpose of the study was to compare the results of the pupils' achievement test scores in the Hawthorne Elementary School, Ottawa, Kansas, for the years, 1939 and 1950.

The final conclusion for the study was, with the achievement test and intelligence test scores for the pupils of 1939 and 1950 as the two basic mediums for the comparisons made, that in a majority of the school subjects tested by the Metropolitan Achievement Tests, the pupils in the Hawthorne Elementary School, Ottawa, Kansas, in 1950 achieved as much as, or in a few subjects, more than did the pupils in the 1939 group. However, where there was found to be a statistically significant difference in mean grade equivalents between the two groups, 1939 and 1950, it was in the majority of cases in favor of the 1939 group.

### Qualifications

The final results were influenced by some factors that were not controlled for the groups under study. The authors recognized factors that would tend to influence the results both in favor of and against the groups of pupils for the two years of the study, 1939 and 1950.

Seven of the teachers in 1950 would be called beginning teachers as they had two or less years of experience, whereas all but one of the teachers for the pupils in the 1939 group had nine or more years of teaching experience.

The majority of the pupils in the 1939 group had not had kindergarten experience as the kindergarten was not started in this school until the 1938-1939 term. The 1950 group had had kindergarten experience.

The 1939 group of pupils were born just before or during the depression years of the early 1930's. Conditions then no doubt had some influence on the emotional growth of the children.

The 1950 group of pupils were born just before, at the beginning, or during World War II. Many fathers were in the service and many mothers worked in defense plants. The environment of these children during the war years could not be compared to those pupils in the 1939 group. There was no way of determining how the emotional dis-

turbances brought on by the war might have influenced the results of the school work for the 1950 group.

#### Recommendations

It is recommended that studies of a similar nature be made in elementary schools in other parts of the country to see how the teaching of the fundamental subjects today compare with that of yesterday.

Administrators and teachers should inform themselves about the results of teaching in order to keep their school public informed and to implement needed improvements in teaching and supervisory practices.

# THE USE OF GRAPHIC REPRESENTATION IN PROBLEM-SOLVING AT THE COLLEGE LEVEL

H. DONALD WHITE  
Extension Division of  
University of Wisconsin

ONE OF THE little realized potentialities of art experience courses for the non-art major student might be that of developing concepts and techniques for problem solving. It is believed that the process of objectifying things, situations, conditions, attitudes, emotions and ideas by projecting symbolical representations of them can aid in clarifying meanings. Mathematics and Geometry and words are the common tools for learning and problem solving, but these means of communication have limitations in usefulness. They are often used to describe pictorial or graphic situations. If the technique of making images and situations more graphic can be accomplished through art experiences a more direct approach to learning and communication can be made.

Psychology would appear to demonstrate that imagery plays a dominant role in our thinking and in the achievement of solutions to problems. As early as 1927 Aveling (1) suggested that images (even at the level of "inner speech") are a condition sine qua non of all our thinking. Comstock (8) tested both the sides of relevancy and irrelevancy of imagery to the process of thought. It was found that imagery was used by all subjects in all problems which included problems in arithmetic, ingenuity, abstractness, general information, completion, enumeration, action, simple judgment, comparison, direction and imagination. It was concluded that there is no irrelevant imagery. Imagery serves in some capacity in any problem even if it serves but to "anchor" the problem-solving approach. Kantor (19) substantiates Aveling in his empirical analysis of meaning, arguing that there is a continuity between what may be called the "perceptual and non-perceptual meaning reactions." He goes on to assert that imagery is a no less definite response of the person than are "meaning-reactions" contained in segments of behavior in which the original object is present. Pear (24:9) argues that abstract images may carry concrete meanings and visa versa. Further, Pear states, "the determining tendencies underlying the abstracting process are all-important, and that whether they eventually issue in a word or as a visual image is comparatively unimportant."

It appears reasonable accordingly to assume that making images graphic can help to objectify any given meaning situation so that solution-achievement can result. However, the rendering of an image requires appropriate symbols to be adopted and any symbol is, of necessity, an abstraction of the original stimulus. A major hypothesis here proposed is that the concepts of abstraction and abstracting in art are closely related to abstract thinking in the general sense. The term "abstraction" is defined to mean, "the process of analyzing and comparing complex objects or situations and determining their common elements, or a generalization so reached."\*\* The term "abstract reasoning" is defined as "the act or process of arriving at conclusions through the use of symbols or generalizations rather than concrete data."\*\*\* These definitions apply equally well to the process of abstract thinking and to abstracting in art creativity. Thus it is proposed that training and experience in representing things and ideas by graphic, abstract symbols will help objectify the elements of the consideration at hand so that learning and solution achievement might result. The forms or symbols represented would depend upon what forms were needed to clarify the meaning-situation. In one case geometrical figures might suffice while in another case symbols of more amorphous character would function to better advantage. The transfer of this training to the various subject areas would be the desired outcome of the experiences. Judd (18:74) writes that, "The very word 'abstraction' signifies a mental reaction in which one neglects, or cuts off, certain impressions in the interest of emphasis on a single line of thought." He goes on to assert that symbolic thinking is economical because it is "a substitute for concrete experience, not a mere repetition or re-statement in memory of such experience." Judd deals at length with mathematical symbols and mathematical abstraction, calling it a process of analysis rather than synthesis. This is where the concept of graphic representation for problem solving as herein proposed departs from the popular concept of abstraction. This is to say that the use of graphic representations of symbols may be, in essence, a synthesizing process in the act of generalizing for solution achievement.

Kurt Lewin's Topological Psychology is a distinctive attempt to represent psychological and sociological considerations of humans by means of graphic representation. More specifically, he attempts to portray the "life space" of an individual by symbols. (21) He admits that there is at present no adequate scientific method for representing life space. His representative system of making topology (the science of spacial relations) graphic is not really an attempt to make abstract

\* Carter V. Good (Editor), Dictionary of Education (New York: McGraw-Hill Book Co., 1945), p. 3.

\*\* Ibid., p. 332.

relationships graphic. He is concerned rather, with the conceptual determination of the dynamic properties of any concrete situation. His system of representative symbols is somewhat limiting since it is, in essence, little more than diagraming.

Hartman (15:134-136) notes that the "doodads" and odd scribblings which men make on table cloths and other surfaces in an effort to make things clear to themselves and others constitute the root of the systematized ideographs of the topologist. Hartmann observes that the drawing of a "picture" of something dimly understood is a way of clarifying meanings. He calls it "behavior portraiture" and labels it as an objectifying experience. Practices like these which involve symbolic characterization tend to facilitate real learning and mastery of the experience, Hartmann concludes.

In an experiment to determine if children could be taught to perform the processes involved in a simple relational syllogism by means of graphic representation, Morgan and Carrington (22) found that they could. Children in grades one to six were used in the investigation. The investigators found that the using of graphic representation by the children aided in their performance. Johnson (17:224) pleads for more use of the fine arts as a means of personality integration. He argues that, "... the creative art process is a process of objectifying all kinds of experiences--those which are consciously before us and those which are stored in memory as vague feelings and impulses." He supports Hartmann in recognizing that when an individual has made something in graphic form he has advanced in understanding something already there. In an effort to bring attention to the unique values in art Dickson (11:242) states, "As soon as we realize that artistic perception is not simply a matter of visual experience a totally new meaning of the development of art is possible."

As a test of the hypothesis that the use of graphic representation can contribute to learning and problem-solving an exploratory investigation was made at the University of Wisconsin in the School of Education during the School Year 1950-51. This investigation was not conceived of as experimental in the strict sense of the word, but rather as one designed to explore possibilities for further investigations. It is to be thought of as a "pilot" study. The study attempted to determine if students trained in techniques of graphic representation and abstract symbolology would show improvement in the abilities of: (a) space relationship perception, (b) problem-solving (logic), and (c) personality improvement (social-adjustment). In addition to ascertaining gains made in these abilities as a result of training, a major concern in the investigation was that of answering the following questions:

1. Can non-art major students learn to represent things, situations, ideas, emotions, and conditions in creative, graphic form?

2. Do learning and increasing problem-solving ability occur as a result?
3. How readily can this problem-solving ability be transferred to other subject areas?
4. Do the students recognize more benefits or uses of art as a result of these experiences?
5. Does an increase in interest in "The Arts" per se occur as a result of experiences of this kind?

Ninety-five students enrolled in a course in the psychology of learning at the University of Wisconsin were selected for the study. This is a course required of all undergraduate Education majors. Of these students, seventy-eight were seniors, eleven were juniors, and six were of graduate standing. Three sections of students were selected as follows: (a) a control group, (b) a lecture group, and (c) a lecture and experience group. Hereafter the groups will be referred to as Group 1 (control group), Group 2 (lecture group) and Group 3 (lecture and experience group). While the total group selected is not a random sample there is no evidence to indicate that it varies greatly from any of the groups enrolled in the Education 75 course at the University of Wisconsin except in details that are not pertinent to the investigation. In any case, there is reason to believe that an investigation of this kind can bring to light many worthwhile tentative generalizations that can later be subjected, under more favorable conditions, to randomized studies. The numbers in the groups were as follows:

Group	Type	Juniors		Seniors		Grad. Students	Total
1	control	6		25		5	36
2	lecture	2		26		1	29
3	lecture and experience	3		27		0	30

### Procedure

The following instruments were administered to the control group and the two experimental groups the week of November 19, 1950: (a) The New California Short Form Test of Mental Maturity, (b) The Space Factor Test of the Chicago Tests of Primary Mental Abilities, (c) The Washburne Social-Adjustment Inventory, and (d) A questionnaire to determine indices of usefulness of art to each student. All of the instruments were administered under prescribed testing conditions to insure objectivity, and all of the standardized instruments are reliable as shown in the manuals of instructions. The tests were machine-scored. The questionnaire requested each student to mark on a scale from one to twenty the degree of usefulness of art to him, and to mark on a similar scale how "The Arts" rated with him as cultural interests.

Training periods were undertaken in Groups 2 and 3. Since transfer of training was the general function desired by the training, care was taken to insure against specific training for the tests. The nature of the training for the groups was as follows:

Group 1--received no training.

Group 2--received two forty minute slide-lectures dealing with the subject, "Art and Abstraction". These lectures were given five days apart. The purposes of the lectures were to:

1. create an awareness of design in our everyday surroundings, and increase visual perception by bringing to consciousness form, color, shape, texture, line and pattern as being implicit in our social and physical environment
2. demonstrate how abstract thought is involved in art creativity
3. reveal the dynamics of abstract symbols for analyzing situations
4. discuss the criteria of good communication.

The two slide-lectures were administered during two regular class periods. Students were encouraged to raise questions and discuss any issue at any point in the presentations.

Group 2 also participated in a class discussion on the subject, "Personality and Social Adjustment." The purposes of this discussion were to:

1. discover what some of the physical, emotional, intellectual, hereditary, socio-economic and environmental influences are on one's "personality,"
2. discover what is essentially involved in social adjustment,
3. ascertain some ways of improving one's social adjustment

Group 3 participated in the same two slide-lectures on "Art and Abstraction," and participated in five art experiences using pencils, crayons and/or watercolor paints on large sheets of paper. These experiences were designed to:

1. teach the expressive possibilities of line, color, shape and texture
2. demonstrate how these "tools" could be combined to render the emotions of joy, anger, sorrow
3. show what is involved in abstracting common elements from natural forms
4. develop an ability to perceive more acutely
5. develop an ability to construct suitable symbolologies for communicating abstract concepts
6. furnish an opportunity to apply the concepts and abilities to individual problems in personal adjustment for purposes of improving social adjustment.

Some of the experience-periods occurred during regular class hours while others occurred at night meetings. The directions were given verbally to the group. As little assistance as possible was given to individual members of the group but occasional aid was necessary, particularly in the technical use of the materials. All of this phase of the training was administered during the two weeks following the initial testing program. Some of the experiences were administered to small groups in several sessions due to the difficulties of scheduling evening meetings for the entire group at one time. The five art experiences undergone by group 3 were as follows:

No. 1 (2 hrs.)

The group was asked to experiment with various color, line, shape, pattern and compositional effects. Then the students were asked to render an abstract (non-realistic) interpretation of the emotions of "joy", "anger" and "sorrow" by the use of compositions of line, color, shape, texture and form.

No. 2 (unlimited time)

The students were instructed to isolate a portion of a desk top and very carefully record the design which was formed thereon by scratches, woodgrain, stains, etc.

No. 3 (2 hours)

The class executed accurate drawings of a number of natural objects such as tree bark, moss, pods, oddly shaped stones, twisted twigs, etc. After these botanically accurate drawings were rendered the students proceeded to extract the essential shapes, textures, colors, etc. from them and to create an abstract composition of the essential elements.

No. 4 (50 min.)

Two situations were presented verbally to the group. The students were then asked to present the situations in graphic form, devising abstract symbols which would best depict the situations.

Situation A

"A weak man and a strong man are locked in a room. Suddenly a hungry lion is thrust into the room. The only other object in the room is a fly swatter."

Situation B

"Picture a large house with a side stripped away. Let it be imagined that many people are variously engaged in that house. In one room a child is playing with a toy, in another room an elderly aunt is resting. In the living-room a gay party is in progress, while the kitchen is a scene of busy preparations for dinner. Now further imagine that the floors and walls of the house dissolve and the action of

each person and activity is frozen in its place."

The second situation was more complex, involving many subtle relationships and requiring greater skill in communication by symbology.

#### No. 5 (unlimited time)

The students were asked to try to look at themselves as if they were two people, the one a diagnostician and the other the subject. They were then asked to identify some of their own personality problems, trying to achieve a high degree of objectivity about the analysis. After the personality problems were identified the students were asked to construct a setting which would give their problems maximal meaning.

The students were then instructed to take their personality problems out of context and, by the use of abstract symbols and the art elements, to portray them graphically much as was done in the last two exercises.

As a separate exercise the students were next instructed to solve the individual personality problems by the use of symbology and by using the art achievements of balance, harmony, rhythm, counterpoint, etc.

None of the students in any of the groups was told at any time of the reasons for performing any of these tasks. The only information given them relative to the investigation was that it was a classroom learning experiment as part of the Education 75 course.

During the week of December 11, 1950 all of the tests and a modification of the questionnaire were re-administered to all the groups.

### Results

An analysis of the test scores was made to discover if the training and experience cited above aided in improving retest scores by groups in significant amounts. This was done by ascertaining the differences between the test and retest scores and by computing mean gains for the groups. Critical ratios were calculated on all the mean gains between groups to determine whether these gains were statistically significant. All critical ratios in the region of statistical significance were then isolated and Students t-test was employed to calculate the statistically significant amount of these differences.

Experimental groups 1, 2, and 3 were broken down into the following subgroups for purposes of making a more discriminating analysis of the effects of the training upon test scores:

Group 1--control

Subgroup 2a--lecture group receiving all three training periods  
Subgroup 2b--lecture group receiving two-thirds of the training as a minimum.

Subgroup 3a--experience group receiving all training.

Subgroup 3b--experience group receiving six-sevenths of the

training as a minimum.

Subgroup 3c--experience group receiving four-sevenths of the training as a minimum.

The time element is an important consideration in an investigation of this sort. It is rather obvious that some things are learned in a few minutes while other material requires weeks, months or years before it is learned. How the element of time affects the results will depend in part upon the nature of the tests employed. As the measuring instruments employed in this investigation are examined, differences in learning-time will be apparent for the different material presented. Accordingly, some of the differences will not show up even though they are potentially present while others may do so, depending upon their complexities. Another factor which would influence test scores is that of low "ceilings" in some of the subtests. This is particularly noticeable with the California S-Form Test of Mental Maturity in which the "ceilings" of some subtests were attained by many of the individuals even in the initial testing. These considerations must be kept in mind in interpreting the findings of this investigation.

The following general statements may serve to reveal the kind of results obtained from the investigation:

1. No significant differences showed up between any of the groups in the spacial relations tests of the California S-Form Test of Mental Maturity nor in the Numerical Reasoning Test involving calculating ability.
2. The control group made the most gain in the vocabulary subtest of the California S-Form Test of Mental Maturity, having gained significantly over both of the lecture subgroups. The experience group with six-sevenths of the training also gained significantly in this ability over the lecture group having two-thirds of the training.
3. No significant differences occurred between any of the groups in the total non-language scores of the California S-Form Test of Mental Maturity, but the control group gained significantly in the total language score over the lecture sub-group which had two-thirds of the training.
4. In the Space-Factor Test of the Chicago Tests of Primary Mental Abilities, both the experience group and the control group gained in statistically significant amounts over the lecture group. The most significant gains were made by the experience group.
5. In a test which involved numerical reasoning of a kind requiring processes of generalizing, recognizing analogies, and making inferences both of the training groups gained significantly over the control group. The most significant gain in this test was made by the experience group.

6. The lecture group gained significantly over both the control and experience groups in the Washburne Social Adjustment Inventory retest. The control group generally made more improvement than the experience group in this inventory.
7. It was found from a subjective analysis of problem solving products that non-art major students can very readily learn to represent situations, ideas, things, emotions and conditions in non-realistic (abstract) form. Evidences of solution-achievement were discovered in these graphic attempts at problem-solving.
8. There is reason to believe that the ability to utilize graphic representation in learning and problem-solving is highly transferable to other subject areas since there is no standardization of problem-solving symbols required.
9. Art was recognized as being more useful by those students who received the lecture and the experience training than it was by the control group. These differences, however, were not in statistically significant amounts.
10. "The Arts" as cultural interests were rated as considerably more important after the training periods by the experience group as compared with the control group.

From the positive findings, it appears that art training of the kinds given to the lecture and experience groups may have had favorable influence on thought processes involving generalization. It also seems probable that art training of the kind requiring keen perceptual abilities may have contributed to skills involving the imaginal manipulation of objects. More time might have aided the experience group. Many cases it would appear that the application of analytical art processes to problem solving may provide an interesting educational departure.

## BIBLIOGRAPHY

1. Aveling, F., "The Relevance of Visual Imagery to the Process of Thinking, II," The British Journal of Psychology, Vol. XVIII, Part I (July, 1927), pp. 15-29.
2. Barlow, M. C., "Transfer of Training in Reasoning" Journal of Educational Psychology, Vol. XXVIII (1937), pp. 122-128.
3. Billings, Marion L., "Problem-Solving in Different Fields of Endeavor," American Journal of Psychology, Vol. XXXXVI (1934) pp. 259-272.
4. Bloom, Benjamin S. and Broder, Lois J., "Problem-Solving Processes of College Students," Supplementary Educational Monographs, No. 73, Chicago: University of Chicago Press, 1950.
5. Brody, Leon, "The Inter-Relations of Verbal Reasoning, Non-Verbal Reasoning, and Intelligence," Journal of Educational Psychology, Vol. XXXI (1940), pp. 295-304.
6. Brownell, William A. and Hendrickson, Gordon, "How Children Learn Information, Concepts, and Generalizations," Forty-Ninth Yearbook of the National Society for the Study of Education, pp. 92-128. Bloomington, Ill.: Public School Publishing Co., 1950.
7. Burack, Benjamin, "The Nature and Efficacy of Methods of Attack on Reasoning Problems," Psychological Monographs, Vol. 64, No. 7 (1950).
8. Comstock, Claire, "On the Relevancy of Imagery to the Processes of Thought," American Journal of Psychology, Vol. XXXII (1921), pp. 196-230.
9. Cunningham, G. Watts, "Perspective and Context in the Meaning Situation," University of California Publications in Philosophy, Vol. XVI, No. 2. Berkeley: University of California Press, 1935.
10. Dewey, John, Art as Experience. New York: Minton, Balch and Co., 1934.
11. Dickson, Elder J., "The Relation of Recent Aesthetic Psychological Research to Art Education," The Journal of Education (London), Vol. LXXIII (June, 1941), pp. 240.
12. Dunker, Karl and Krechevsky, I., "On Solution-Achievement," Psychological Review, Vol. XXXVI (1939), pp. 176-185.
13. Fisher, Sara C., "The Process of Generalizing Abstraction; and its Product, the General Concept," Psychological Monographs, Vol. XXI, No. 2 (1916).
14. Griffitts, Charles H., "Individual Differences in Imagery," Psychological Monographs, Vol. XXXVII, No. 3 (1927).
15. Hartmann, George W., Educational Psychology. New York: American Book Co., 1941.
16. Heidbreder, E. F., "Problem-Solving in Children and Adults," Pedagogical Seminary and Journal of Genetic Psychology, Vol. XXXV (1928), pp. 522-544.

17. Johnson, R. E., "Fine Arts as a Means of Personality Integration," School Review, Vol. LVI (April, 1948), pp. 223-228.
18. Judd, Charles H., with the cooperation of Breslich, E. R., Mc Callister, J. M. and Tyler, Ralph, Education as Cultivation of the Higher Mental Processes. New York: The MacMillan Co., 1936.
19. Kantor, J. R., "An Objective Interpretation of Meanings," American Journal of Psychology, Vol. XXXII (1928), pp. 231-248.
20. Krawiec, T. S., "Comparison of Learning and Retention of Materials Presented Visually and Auditorily," Journal of General Psychology, Vol. XXIV (April, 1946), pp. 179-195.
21. Lewin, Kurt (translated by Fritz and Grace Heider), Principles of Topological Psychology. New York: McGraw-Hill Book Co., 1936.
22. Morgan, John J. B. and Carrington, Dorothy H., "Graphic Instruction in Relational Reasoning," Journal of Educational Psychology, Vol. XXXIII (1944), pp. 536-543.
23. Ogden, C. K. and Richards, I. A., The Meaning of Meaning. New York: Harcourt, Brace and Company, Inc., 1945.
24. Pear, T. H., "The Relevance of Visual Imagery to the Process of Thinking, I," The British Journal of Psychology, Vol. XVIII, Part I (July, 1927), pp. 1-14.
25. Smoke, Kenneth L., "The Experimental Approach to Concept Learning," Psychological Review, Vol. XXXXII (1935), pp. 274-279.

# TEACHERS' EDUCATIONAL BELIEFS VERSUS THEIR CLASSROOM PRACTICES

W. A. OLIVER  
Portland Public Schools  
Portland, Oregon

THIS ARTICLE IS based upon a study of the problem of educational lag, as represented by a study of the stated educational beliefs of one hundred nineteen elementary teachers, as contrasted to their classroom practices. A great deal has been written on the problem of education lag and the fact, that in general, classroom practices of most teachers are not consistent with their stated educational beliefs or with recognized principles of learning. The following quotations offer documentary evidence of this viewpoint. Dr. Hollis Caswell of Teachers College and Dean J. Murry Lee of the State College of Washington state as follows:

"The gap between the best practice and average practice in elementary education is very great, and between the best and the poorest it is tremendous." <sup>1</sup>

"Teachers render lip-service to the provisions for individual differences, but many of their teaching procedures carefully avoid them." <sup>2</sup>

These statements are particularly significant, as it is generally agreed that the greatest progress is adaptation of teaching methods to research has taken place in the elementary school.

This study was geared to four principles of learning which have gained universal acceptance among educators during the past quarter century. The four principles selected are as follows:

1. Good teaching recognizes and provides for individual differences among children.
2. Human growth and development is a continuous process.

1. Hollis Caswell, Education in the Elementary School (New York: American Book Co., 1942), p. 2.

2. Dorris Lee and J. M. Lee, The Child and His Curriculum (New York: D. Appleton-Century Co., 1950), p. 168.

## CLASSROOM EVALUATION REPORT

Please fill out this report as soon after visiting the classroom as possible, by checking in the column which most adequately describes the learning situation.

Teacher Number \_\_\_\_\_ Grade Taught \_\_\_\_\_  
Dates of Observation \_\_\_\_\_

- I. INDIVIDUAL DIFFERENCES
    - A. Was there evidence of differentiated assignments?
    - D. Were any tests used to determine pupil needs?
  - II. HUMAN GROWTH AND DEVELOPMENT IS CONTINUOUS
    - A. Were subject areas integrated into a series of learning experiences suited to the maturity level of the learners?
  - III. REAL LEARNING IS EXPERIENCING
    - A. Were any units developed by the pupils with the guidance of the teacher?
  - IV. LEARNING PROCEEDS BEST WHEN RELATED TO THE INTERESTS AND EXPERIENCES OF THE LEARNER
    - A. Was there evidence of using pupil interest and experience in planning and selecting learning activities?
    - C. Were Arithmetic and Language taught in a formal situation unrelated to the daily life experiences of the pupils?

3. Real learning is based upon experiencing.
4. Learning proceeds best when related to the interests and experiences of the learner.

Any review of the educational literature of the last twenty-five years will easily establish the universality of these four principles. They are referred to in book after book dealing with learning and the learning process. John Dewey through his writings and pronouncements did much to promote their acceptance. They have also been reiterated by such recognized authorities as Burton, Cole, Lee, Hildreth, Spears and Witty.<sup>3</sup> These are but a few of the leaders and writers in the field of education, who continually refer to the above principles as basic to an effective program of classroom instruction.

This study was conducted in six elementary schools in a large American city. The schools were typical elementary schools representing a cross-section of the community. They were selected on the basis of the willingness of the staffs of each school to participate in such a study. There were one hundred nineteen teachers representing all grade levels from kindergarten through the eighth grade, since the seventh and eighth grades in the city where the study was conducted are a part of the elementary schools. The following table gives the statistics upon the one hundred nineteen teachers who participated in the study. It was felt that this was an adequate sampling to indicate whether the problem of educational lag is a serious one.

These statistics indicate that the group studied was a representative cross-section of the elementary teacher population. It involved teachers of all age levels, experience, training, and tenure status.

To determine the extent to which these teachers accepted the concepts embodied in the four principles listed previously, a fifty-item check list of educational beliefs was prepared. The items were taken,

<sup>3</sup> W. H. Burton, The Guidance of Learning Activities (New York: D. Appleton-Century Co., 1940).

Luella Cole, Teaching in the Elementary School (New York: Rinehart Publishing Co., 1946).

G. Hildreth, Child Growth Through Education (New York: The Ronald Press, 1946).

Dorris Lee and J. M. Lee, The Child and His Curriculum (New York: D. Appleton-Century Co., 1950).

Harold Spears, Some Principles of Teaching (New York: Prentice-Hall Co., 1949).

P. Witty and C. E. Skinner, Mental Hygiene in Modern Education (New York: Farrar and Rinehart, 1941).

TABLE I

SHOWING STATISTICS ON THE ONE HUNDRED NINETEEN ELEMENTARY  
SCHOOL TEACHERS STUDIED

Median Age of All Teachers	42.2
Median Number of Years Taught	9.0
Number of Teachers with Master Degrees	6.0
Number of Teachers with Bachelor Degrees	65.0
Number of Teachers without a Degree	48.0
Number Kindergarten-Primary Teachers	52.0
Number Teachers in Grades 4, 5, and 6	39.0
Number Teachers in Grades 7 and 8	28.0
Number of Teachers on Tenure	71.0
Number of Teachers not on Tenure	48.0

for the most part, from two sources—one list having been prepared by Dr. Mort and Dr. Connell<sup>4</sup> of Teachers College, and on by Dr. Spalding<sup>5</sup> of the University of Illinois.

These statements were aimed at determining the extent to which the teachers involved in the study agreed or disagreed with the concepts embodied in the statements. Their responses would also give a clue to the degree of acceptance or rejection of the four principles underlying the study. The following statements are samples of the type of statement used. The number indicates the placement of the statement in the series.

3. Each child is different from every other child and the school should treat each child as an individual.
4. Success in learning is greatest when learning results from the natural incentives and interests of the child.
14. Textbook learning should give way to increased learning through experience.
26. It is more important for a teacher to have a thorough knowledge of subject matter than to know how children learn.
35. A pupil acquires self-discipline only through freedom to make choices and to profit from them.
50. The curriculum should consist of worthwhile activities on the child's own level.

The teachers indicated their acceptance or rejection of these statements by checking their responses as "agree" or "disagree." Table II shows the results of this phase of the study.

These scores show a high degree of acceptance of the stated beliefs on the part of the teachers. This would also indicate a high degree of acceptance of the concepts embodied in the four principles of learning selected as basic to a good educational program. Seventy-five per cent of the teachers agreed with eighty per cent of the statements and only two individuals failed to register agreement with less than half of the items. This speaks well for the writings and teachings of the educational authorities over the past quarter century. The philosophy of modern education has been well accepted, but what of the implementation of it?

A further analysis of these scores in relation to the factors of age,

4. F. G. Connell and Paul Mort, A Guide for Self-Appraisal of School Systems (New York: Bureau of Publications, Teachers College, Columbia University).

5. W. B. Spalding, Survey of Teachers' Educational Beliefs, Official Records, School District No. 1, Portland, Oregon.

TABLE II

SCORES MADE BY 119 TEACHERS ON  
STATEMENTS OF EDUCATIONAL BELIEFS

Score	Frequency	Score	Frequency	Score	Frequency
50	1	42	8	35	3
49	6	41	6	34	1
48	10	40	6	33	3
47	8	39	9	32	3
46	15	38	3	31	1
45	7	37	8	30	2
44	11	36	2	23	1
43	6				
Median	43.3				
Q-1	38.8				
Q-3	46.1				

training, experience, grade level taught and tenure failed to develop any significant relationships. That is, acceptance of these beliefs was not peculiar to any single group within the one hundred nineteen teachers.

The second phase of the study was concerned with the evaluation of the classroom practices of the one hundred nineteen teachers as reflected in the learning situations in their classrooms. The first task was the development of an evaluative device geared to the four recognized principles of learning stated at the beginning of the study. After considerable research and discussion, and with the assistance of the Department of Research, a suitable device was prepared. The instrument finally agreed upon consisted of four main divisions, each one representative of one of the four principles of learning on which the study was based. Under each of these appeared a number of sub-heads descriptive of types of educational practices one would expect to find in a classroom implementing the stated principles. The following is a partial reproduction of this device.

A total of fourteen such statements were included in this evaluative device. It can be seen that the attempt was not to rate the teacher, but to determine to what extent the classroom situation reflected the concepts embodied in statements of educational beliefs. The following table shows the results of the evaluation.

A perfect score on evaluation would have been 52. It is interesting to note that seventy-five per cent of the group scores 20.8 points below this level and that the highest score was 44. From these scores it is evident that there is little relationship between the professed educational beliefs of these teachers and their classroom practices. The correlation between the belief scores and the evaluation scores was .31, an extremely low correlation indicating practically no relationship. This is not a condemnation of the classroom teacher, but rather indicates the failure to provide teachers with the basic understandings and techniques necessary to provide the type of learning environment consistent with our accepted principles of learning. The following conclusions are inevitable as a result of this study.

1. The problem of educational lag is a serious one in education.
2. Teachers in general have little real understanding of the basic principles of child growth and development.
3. Teachers have not been given the necessary techniques to develop a classroom program based on child needs, interests, and capacities.
4. The actual provision for individual differences in most classrooms is very limited.
5. The learning experiences are, in many cases, still limited to assignment-recitation type of activity.

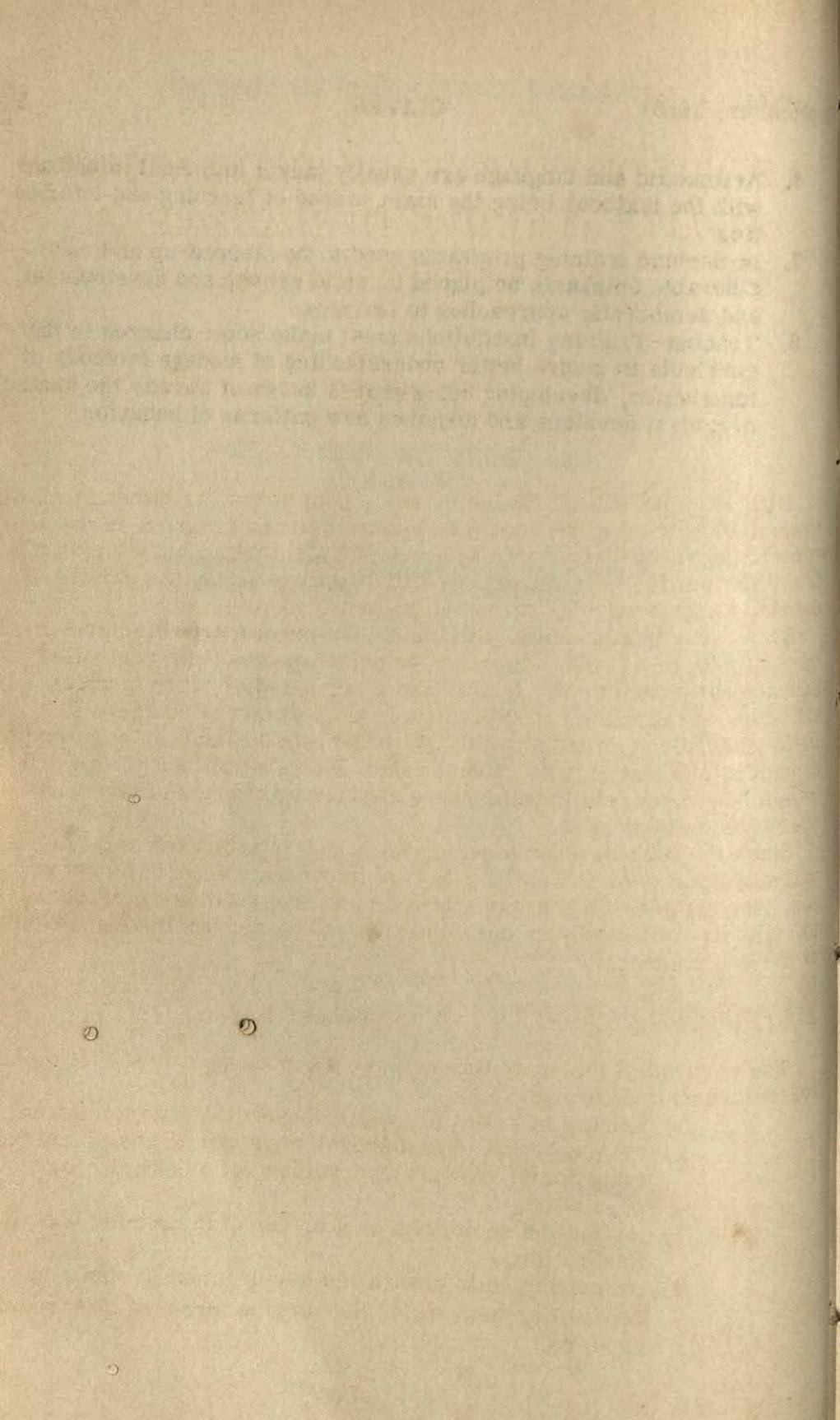
TABLE III  
EVALUATION SCORES OF 119  
TEACHERS

Score Interval	Frequency
43-45	3
40-42	3
37-39	7
34-36	11
31-33	10
28-30	29
25-27	29
22-24	17
19-21	6
16-18	3
13-15	1

(2)	Median	27.8
	Q-1	24.7
	Q-3	31.2

6. Arithmetic and language are usually taught in formal situations with the textbook being the main source of learning and information.
7. In-Service training programs need to be stepped-up and considerable emphasis be placed on child growth and development, and democratic approaches to learning.
8. Teacher-Training institutions must make some changes in their curricula to insure better understanding of modern methods of instruction, developing out of what is known of the way the human organism develops and acquires new patterns of behavior.



# A STUDY OF PEER GROUP SOCIAL ACCEPTABILITY AT THE FIFTH GRADE LEVEL IN A PUBLIC SCHOOL

ELVA E. KINNEY  
Greenville College  
Greenville, Illinois

DURING THE MIDDLE elementary school years, as the child moves toward adolescence, he should be making marked progress in the accomplishment of that important developmental task, - emancipation from the family. Simultaneously with progress along that line there develops a growing importance of the peer group.

This study is concerned with the child's normal growth in acceptability with this group. It makes no pretense of being a controlled educational experiment. It is rather a comparative study in which all seven of the fifth grade teachers in a small city in southern Illinois worked cooperatively with the writer as consultant in an attempt to understand and improve social relationships within the groups and to see how their relationships were affected by the type of group in which the child worked.

Since the time devoted to supervision had to be limited to three half-days per week it was thought best to put the emphasis on reading. Nevertheless there was a tendency where possible for teachers to make use of techniques developed in reading through the remainder of the day's program.

## The Purpose of the Study

The purpose of the study was to make a contribution to the field of teacher education through

- 1) helping teachers to a better understanding of children,
- 2) helping teachers to discover ways and means of objectifying social growth and development in school room population,
- 3) testing the sociogram as a means of measuring social integration,
- 4) comparing child growth and development in social acceptability made under the various forms of classroom grouping,

## CHART I

PATTERN OF THE SOCIOGRAM MADE FOR THE  
DIFFERENT GRADE GROUPS OF THIS STUDY

These numbers indicated the frequency of choices given each individual

---

12  
11  
10  
9  
8  
7  
6  
5  
4  
3  
2  
1

Each girl was indicated by name in a circle on the left side placed at the level which showed the number of choices she received.

Each boy was indicated by name in a square on the right side at the level which showed the number of choices he received.

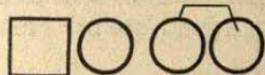
Lines from "chooser" to "chosen", projected one-eighth inch within the figure representing the one chosen, indicated social relationships within the group.



September, 1953)

KINNEY

59



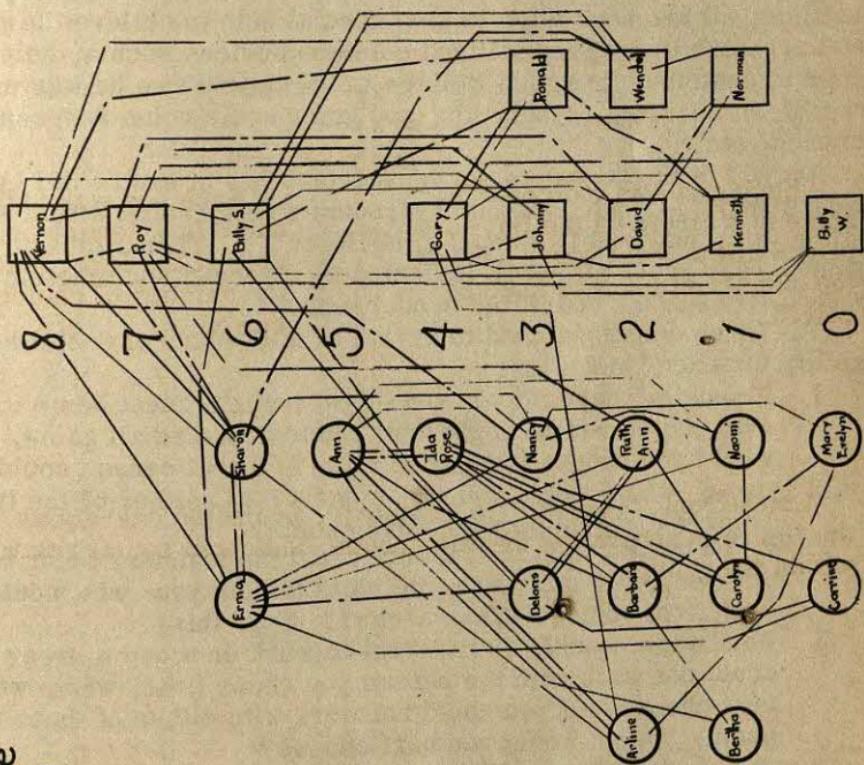
First Choice  
Second Choice  
Third Choice

Boy

Girl

One Choosing

One Chosen



School No.3  
Guest in Home  
December 1

Chart No.13

- 5) pointing the importance of training teachers in analyzing a social group and building sound social relationships.

### Procedures

Three forms of grouping were used:

- 1) the entire grade group,
- 2) ability groups,
- 3) the grade group supplemented by frequent use of small, informal, flexible groups.

Each teacher chose to work toward the usual academic goals of the school through the medium of one of these three types of grouping. In addition, all teachers tried to give special help to children low in social acceptability. They utilized common devices such as helping the child to contribute to group welfare in the field where he was most capable of securing recognition, assigning some room responsibility to each, etc.

A study of the degree of social acceptability of each child by his peer group was made, and interpersonal relationships within each group were analyzed by means of sociograms. Then a measure of gain in peer group status as indicated by a second sociometric study was made near the end of the school year.

The three questions used to arrive at a measure of social acceptability for each child were:

1. If your mother told you you might invite a guest home to dinner with you on Friday night and then to a basketball game, whom would you choose from your room? If that person couldn't come, whom would you choose next? If neither of the two could come who would be your third choice?
2. If you needed help in arithmetic and the teacher, being busy, told you you might ask some child to help you, who would be your first choice? Your second? Your third?
3. With whom would you like best to work on a committee? If you could not work with the person you chose first, whom would you next choose? If you could not work with either of these two people, whom would you next choose?

A special adaptation of Moreno's sociogram was used to record these choices. A pattern of the form used is shown on Chart I.

In studying these sociograms statistically a first choice was weighted as six points, a second choice as three points, and a third choice as one point. Since according to this schedule 30 points would indicate a statistically average acceptability score, (3 firsts, 3 seconds, and 3 thirds) a score one third of this (10 points) was accepted as the critical score below which the child was considered in need of help in social acceptability. Children with scores of 0 were designated as isolates.

TABLE I

NUMBER OF ISOLATES AND FRINGERS IN EACH OF SEVEN FIFTH  
GRADES USING VARYING GROUP PROCEDURES THROUGHOUT THE YEAR

School	Number Enrolled	Median I. Q.	Range of I. Q.	Type of Grouping for Classroom Activity	Number with Social Acceptability Score Below 10				
					December 1			April 20	
					Isolates	Fringers	Total	Score 0-9	Score 1-9
1	32	110	73-145	Entire grade	3	4	7	4	5
2	27	102	78-127	Entire grade	0	6	6	2	6
3	26	104	73-128	Small Flexible groups	0	7	7	0	5
4	30	110	76-139	Small flexible groups	1	8	9	1	5
5	28	89	63-115	Ability groups	2	6	8	2	7
6	32	103	72-140	Small flexible groups	2	7	9	0	6
7	34	104	73-126	Ability groups	1	9	10	1	12
									13

The percentage of isolates and fringers of total enrollment in schools using flexible grouping was 28.4 in December and 19.3 in April. In the other groups it was 23.9 in December and 32.2 in April.

and those with scores between 0 and 9 as fringers.

Sociogram results were also used as a measure of group cohesiveness. Except for the nature of the grouping, the same techniques were used in all fifth grades in an attempt to bring about integration of the grade groups.

Intelligence tests at the beginning of the year were used to give one more clue to a better understanding of the boys and girls. Also, achievement tests were given at the beginning and the end of the year, the first to help in appraising assets and liabilities of each child and the two together to get a picture of academic gain under each method of grouping.

Throughout the year the writer as unofficial supervisor was available for consultation with single teachers or groups of teachers, for actual development with the children of new procedures, for help in finding materials, etc. Three half-days per week or their equivalent were spent in these activities or in observing the groups at work.

## Results

### Comparative Social Changes in the Different Fifth Grades.

This investigation showed that in the population studied it was possible in five months time to increase significantly the social acceptability of children who at the beginning of the study were not socially an integral part of their grade group.

Table I compares the number of isolates and fringers at the time the first sociometric test was given with the number at the time of the second. In schools 3, 4, and 6, the total number of isolates had been reduced from three to one and the fringers from twenty-two to seventeen. It will be noted that in grades where small informal flexible groups had been frequently used the number of isolates and fringers had in every case decreased. In the remaining four groups the number of children with low acceptability scores had without exception increased during the period of the study. This is true of each individual group as well as of the total. Isolates had increased from six to nine and fringers from twenty-five to thirty.

This suggests the possibility that there may be a factor of importance in raising the social acceptability of children that teachers personally with all their best efforts are unable to affect except as they may create and maintain the environment to produce the results. In other words, it may be that these desirable changes in degree of social acceptability can only be brought about as children have an opportunity to react in a group of their peers of such size that there can be pleasant successful social interaction. In fact, it may be necessary to plan first of all for such interaction between two children only and then progress gradually toward the larger group.

The higher per cent of children with social acceptability score in

the middle range, in grades where small flexible grouping had been frequently used, indicated that this device helped not only to raise social acceptability scores but also to create more cohesiveness in the entire grade group.

#### Comparative Academic Gains in the Different Fifth Grades.

While there was a very slight difference in achievement in favor of the grades using small groups, other factors made it impossible to claim a distinct advantage for these groups.

#### Factors that Helped to Make Supervision Effective.

While there is no statistical evidence as to the factors in this investigation that helped to make supervision effective, it is the opinion of the writer, after conferring with administrators, principals, and teachers that the following were recognized by all concerned as important:

1. The whole-hearted cooperation of the administration.
2. The maintenance of a cooperative rather than a dictatorial relationship between teachers and supervisor.
3. The fact that the supervisor made a thorough study of social and economic background in the various schools as a basis for the study.
4. The fact that the sociogram, as used in this study, objectified the work in social relationships.
5. The fact recognized by the teachers that the supervisor, or consultant, had nothing to do with their ratings, salary, hiring or status with the administration. They felt free to express their real thoughts and emotions. For curriculum development, it seems important for the teachers to have association with someone in whom they have confidence--someone who will provide for their basic human needs: security, recognition, friendliness, success and freedom.

#### Implications of the Study

It is believed that this study has important implications for supervisors, for administrators, and for classroom teachers. The following statements are not statistically proved in this study. They are deductions which the writer believes to be important, and for which she believes there was anecdotal evidence. The principal evidence of such concepts originated in the experience of meeting with teachers individually and in groups. Also the many conferences with administrative and supervisory personnel influenced the judgments expressed below.

Implications for Supervisors.

1. Supervision is most successful when it is actually educational leadership in cooperative planning.
2. Teachers have a need of leadership in a program aimed to meet social acceptability needs of children.
3. The supervisor should meet the basic needs of teachers; recognition, security, etc.

Implications for Administrators.

1. If administrators hope to see maximum teacher growth as a result of supervisory activities, supervision must not be tied up with ratings, salary, hiring, etc.
2. In making plans for effective supervision, consultant service needs to be emphasized.

Implications for Teacher Education.

1. Teachers need to acquire techniques for meeting social acceptability needs of children. Especially they need to discover the role of small, informal, flexible groups in meeting those needs.
2. Teachers need to discover techniques for developing group cohesiveness.
3. The sociogram as a method of appraising social acceptability needs and measuring gains in that area is a tool which classroom teachers need to learn to use.

Suggestions for Further Study

1. Long-time studies of groups and their social relationships need to be made to discover whether or not these gains in social acceptability are permanent.
2. Long-time studies also need to be made to see how far it is possible to reduce the number of isolates and fringers. In other words, with sufficient time and effort is it possible to help all children to attain a degree of social acceptability?
3. We need many studies of the use of sociometric grouping in all parts of the country to determine how general the conclusions of such a study as this are.
4. We need studies of techniques for awareness of group dynamics.

# A COMPARISON OF THE EXTENT OF RETENTION ACCOMPANYING THE USE OF THREE TEACHING-TESTING PROCEDURES

LLOYD L. PATHAEL  
E. D. FITZPATRICK  
LEDFORD J. BISCHOF  
Southern Illinois University  
Carbondale, Illinois

## Purpose

THE PURPOSE OF this study is to determine whether the use of the SRA<sup>1</sup> Self-Scorer contributes to greater retention of course material than other methods of teaching-testing. It has been claimed that the self-scoring by yielding immediate knowledge of the correct response to a test item brings about greater retention. The basic assumption for this claim seems to be that when the student is motivated to a high degree toward a goal, the satisfaction that results from reaching the goal rather quickly should result in retention of the satisfactory response. This device also seems to be based on the assumption that the rate of forgetting is most rapid immediately after the response as illustrated by the Ebbinghaus Curve.

The self-scoring is so constructed that when the student punches out the circle representing the correct response the underside of the pad will show through in red. If he punches out a circle representing the incorrect response, the underside of the pad will show through in white and the student may then take a second, third, or even fourth choice until the red shows through.

In the first instance the student is in a highly competitive situation and is generally striving for the correct answer to the test questions. On the SRA Self-Scorer the student is rewarded rather quickly with the knowledge of the correct response. In the second instance he has less time to forget the particular test question, his answer, and the correct response than if he were answering the questions on another device. However only by experimental results can these claims be proven or disproven.

1. Science Research Associates.

TABLE I

MEAN PERFORMANCE OF FOUR CLASSES BY METHODS  
OF TEACHING-TESTING WITH TOTALS

Class	Methods			Totals	Means
	SRA	IBM	New Questions		
1	35.08	40.21	31.68	106.97	35.66
2	31.84	32.54	28.16	92.54	30.85
3	38.63	39.24	31.71	109.58	36.53
4	32.30	34.15	28.76	95.21	31.74
Totals	137.85	146.14	120.31	404.30	
Means	34.46	36.54	30.08		33.69

### Method and Results

Four classes in psychology totaling 146 pupils largely freshmen and sophomores in the College of Education, Southern Illinois University, Carbondale, Illinois, were used in this investigation. Two classes were taught by one teacher; two were taught by another. Through twelve weeks of instruction the teaching procedures were determined by the instructors with this exception. Each test during the twelve week period was a combination of two procedures. One half of the test was made up of questions to be answered on the SRA Self-Scorer while the other half of the test consisted of different questions to be answered on the IBM machine scored answer sheets. The questions were alternated for each of his classes by each teacher, so that one class answered one group of questions one way and the other class answered the same questions the other way thereby removing variance due to differences in difficulty of the test questions. Following each examination the questions answered using the SRA Self-Scorer were not discussed. Five such tests were given during the term. The final examination was then constructed using a selection of items that had been answered on previous tests by means of the self-scoring for the first part, a selection of items that had been answered on the IBM sheet for the second part; and completely new test questions for the third part. The test questions selected were questions from previous tests that seemed to be the most valid for the course content and showed the best discriminative powers. The answers to these questions were re-keyed to eliminate the retention of symbol response. An analysis of variance was applied to the data collected from the final examination on the basis of the null hypothesis that the observed differences between mean performance of the four classes on the three parts of the examination were due to chance errors and that no real difference existed.

Table I illustrates the mean performance of the classes by methods. By determining the variance ascribed to the various classes, and by determining the variance due to the methods the amount of the total variance that remains can be shown after these two factors are removed. The remaining variance then becomes the error estimate and takes into consideration all sources of error that have been randomized within each class.

Table II illustrates the results of the analysis. The ratio between the remainder variance and the methods variance is compared to the ratio expected on the basis of normal probability. The ratio in this case is considerably larger than that expected at the 1% level of significance and disproves the null hypothesis. However to evaluate the individual differences in methods means further computations are necessary.

TABLE II  
ANALYSIS OF VARIANCE

Type of Variance	df	Sum of Squares	Variance	F ratio
Methods	2	87.637	43.82	
Classes	3	72.117	24.04	
Remainder	6	10.303	1.72	
	2 and 6			25.4767

P lies beyond .01

TABLE III  
SIGNIFICANCE OF DIFFERENCES BETWEEN PAIRS OF MEANS  
AT THE 1% LEVEL

Pairs of Means	Differences	Significant at 1% Level
SRA-New Questions	4.38	Yes
IBM-SRA	2.08	Yes
IBM-New Questions	6.46	Yes

The standard error of the difference between means for any two methods was calculated as well as the maximum error in score units at the 1% level. The standard error of the difference amounted to .153 therefore the critical difference between means at the 1% level in score units would be .567, as the t for 6 degrees of freedom is 3.707.

Table III illustrates the significance of difference between pairs of means. The IBM teaching-testing technique shows marked superiority over the SRA technique and the new questions technique. The SRA technique is markedly superior to the new questions technique.

### Conclusions

From the results obtained certain conclusions can be drawn.

1. Testing as a part of teaching method contributes to the retention of course material.
2. The discussion of test questions answered on the IBM answer sheets and their correct responses following an examination contributes to greater retention of course material than testing with the SRA Self-Scorer without discussion or no testing at all(represented by questions completely new to the student).
3. Testing with the SRA Self-Scorer contributes to greater retention than no testing at all.

Further research is necessary to determine whether the IBM technique would still show superiority over the SRA technique if a discussion of questions also followed the use of the SRA technique.

# CORRELATIONS MEANING, THROUGH KNOWN CO-VARIANT COMMONALITIES\*

HENRY F. DICKENSON  
Lincoln Memorial University  
Harrogate, Tennessee

TO ELIMINATE possible errors in prior coin tossing experiments, a new coin dropping apparatus was constructed consisting of three perpendicular coin slots above each of four separate compartments. The partitions to these compartments did not extend to the floor of the apparatus. This enabled a thin periodical to be slipped under the compartments to reduce the noise of coin drops. A floor plan of the four compartments (I, II, III and IV) was drawn on the periodical. After each complete coin drop it was pulled forward on the extended floor of the apparatus for easy access to the coins of all four compartments.

The collection of data for the experiment consisted of taking eight pennies in each hand and dropping four, (one at a time) into each compartment. One was dropped at a time in order to prevent interference of one with another in falling. A record was kept of the number of heads dropped in each of the compartment I, II, III, and IV for 4,034 complete drops. Later needs, however, necessitated eight different compartmental drop columns with four pennies to each. And since compartmental co-varient comparisons were possible either horizontally or vertically, this actually doubled the compartmental drops of 17,136 individual pennies to 34,272 individual penny drop possibilities.

It was thought that coins dropped on edge, one at a time in one compartment, would yield a chance order of heads with any other compartment and a corresponding zero correlation. This contention was confirmed by correlating 5,191 pairs of adjacent compartmental drops. The correlation coefficient by the product-movement method was  $r = .00$ , indicating chance order, or no relationships.

With zero relationship established between the number of heads in adjacent compartments, known amounts of perfect co-variance could now be studied effectively. The following table summarizes the results. Four thousand pairs of  $x$  and  $y$  compartmental combinations were considered to constitute a minimum universe in each category

\*This paper is the concluding one of a series based upon more than a decade of experimenting with upwards of fifty thousand co-varient relationships.

below. Twenty-five percent  $x$  and  $y$  commonality was obtained by combining three non-overlapping and one overlapping compartment.

A critical ratio of at least 4.00 is necessary for a statistically reliable difference between means (5). There is no reliable difference between mean co-variant commonality and actual numerical correlation coefficient.

Table I summary, with no co-variance commonality, resulted in a coefficient of correlation having zero numerical value. Table II summary, with an average twenty-five percent co-variance commonality, resulted in a coefficient correlation having twenty-three hundredths numerical value.

Table III summary with an average co-variance commonality of fifty percent, resulted in a coefficient of correlation having a numerical value of fifty hundredths. Table IV summary, with an average of seventy-five percent co-variance commonality resulted in a coefficient of correlation having a numerical value of seventy-five hundredths. Therefore average co-variance commonality and correlation magnitude are the same. Historically, (5, 9, 10, 2) this is true as well as from the inductive conclusion drawn from the first five summaries tables, and from the 23,184 cases above.

The next five table summaries derived from 27,184 cases show just as conclusively that the correlation coefficient and the size of the mean co-variance commonality is the same. Since the mean co-variance commonality is derived by simple arithmetic. Example,  $r$  equals .75, the mean co-variance commonality is likewise .75. Then if  $y$  axis has a known co-variance commonality of .50,  $x$  co-variance commonality, as represented in the last half of the table summaries, as well as they did in the first half. There has been no end to the confusion emanating from such correlation interpretation (6, 7, 1, 11).

Such meanings of correlation coefficients, as interpreted by statistical formulas, were derived originally from this type of error (4). Namely, that of calling  $r$  equals .707 or  $r$  equals .866 fifty percent of a perfect correlation.

These findings make it increasingly imperative that one standard become the basis for interpretations. There is no better basis than knowing the mean  $x$  and  $y$  co-variance commonality, and accepting the resulting correlation coefficient for what it is—the same degree of perfect relationship. Instead of complexity and confusion it appears that the coefficient of correlation should be retained as a general concept, signifying numerical degree of perfect direct or inverse relationship. In reality, it is as simple as a percent relationship, and should be thought of as a percentage of perfect co-variance, when the minimum universe is represented.

TABLE I  
KNOWN CO-VARIANCE COMMONALITY AND ACTUAL CORRELATIONS

Table (No)	Cases (n)	x %	y %	Commonality	m %	Actual (r)
I	5,191	.00	.00	.00	.00	.00
II	4,546	.25	.25	.25	.23	
III	4,086	.50	.50	.50	.50	
IV	4,307	.52	.52	.51	.52	
V	5,054	.75	.75	.75	.75	
VI	5,940	1.00	1.00	.50	.75	.75
VII	7,707	1.00	1.00	.50	.75	.75
VIII	4,000	1.00	1.00	.50	.75	.75
IX	4,000	1.00	1.00	.75	.88	.87
X	5,543	1.000	.80	.90	.90	
Total	50,369	70	53	M = .613	M = .607	

Critical Ratio (5) equals .85

Implications

1. When a universe is represented with adequate data, either high, low or intermediate, coefficients of correlation are equally valid.
2. The simplest explanation of a correlation would appear to be the best one.
3. When one knows that .866 percent of the data in a universe are perfectly correlated, it stretches the imagination to say that this is only half way between no relationship,  $r$  equals .00, and perfect relationship,  $r$  equals 1.00.
4. Chance relationship represented by  $r$ , is not  $r$  equals .50 as is sometimes supposed, but  $r$  equals .00.
5. For prediction purposes, a correlation of  $r$  equals .15 may be just as valid as an  $r$  equals .75, if it represents enough cases to be reliable.
6. Fluctuations in the amount of commonality due to improper coin tossing may lower or raise the correlation.
7. The coefficient of alienation has served its purpose and would therefore do well to join the Ptolemaic rotation theory of the sun.
8. The correlation coefficient formula designed to agree with the concept of coefficient of alienation (so called one half overlapping of one axis and the totally ignored overlapping of the other axis) belongs also in the category of 7 above.
9. The relativity of the universe makes the correlation concept of utmost importance for its comprehension.
10. Finally the correlation concept should be improved by simplification, for a wider utility in revealing knowledge, not concealing it.

## BIBLIOGRAPHY

1. Brown, Wm., Essentials of Mental Measurements, 1925, p. 141
2. Darbishire, A. D., Some Tables for Illustrating Statistical Correlation Mem and Proc. Manchester Lit. and Phil. Soc., Vol 51, pp (of reprint) 1-20, 1907
3. Ezekiel, M., Methods of Correlation, 1930, p. 120
4. Dickenson, H. F., Primary and Secondary Discriminative Reactions, George Peabody College, Nashville, Tennessee, 1933
5. Garrett, H. E., Statistics in Psychology and Education, 1937, pp 352-354
6. Hull, C. L., "The Joint Yields from Teams of Tests", Journal of Educational Psychology, 1923, 14, pp 396-406
7. Kelly, T. L., "Principles Underlying the Classification of Men", Journal of Applied Psychology, 1919, pp 50-67
8. Monroe, W. S., "Note on the Interpretation of the Coefficient of Correlation", Journal of Educational Psychology, 1936, 27, pp 551-553
9. Pearl, R., Medical Biometry and Statistics, 1930, p 370
10. Pearl, R., Medical Biometry and Statistics, 1940, p 302
11. Tryon, R. C., "The Interpretation of the Correlation Coefficient", Psychological Review, 1929, 36, pp 419-445
12. Walker, Helen M., Elementary Statistics, 1944, pp 203-205, 357

## BOOK REVIEWS

L. Joseph Lins, University of Wisconsin; Valworth Plum, University of Minnesota;  
John Schmid, Michigan State College

Arny, Clara Brown. The Effectiveness of the High School Program in Home Economics. (Minneapolis: University of Minnesota Press, 1952), pp. xvii + 319, \$4.75.

An experimental project conducted under the direction of the author in twenty representative Minnesota high schools is reported in this book. Believing that two common weaknesses in educational research are that "it is continued over too short a period of time and that too few aspects of the situation are studied", the director undertook a five year investigation during which time many techniques and procedures were tried out and appraised in terms of usefulness in picturing various aspects of the home economics program and student performance.

It was believed that such a project could demonstrate the fact that educational research can serve to locate strong and weak points in a program and can identify factors influencing effectiveness and means successful in bringing about improvement. A core of research assistants participated in the collection and analysis of a wide variety of data. An advisory committee with representatives from the State Department of Education, the University of Minnesota and the participating schools, functioned throughout the study.

After discussion of the general nature of the project, details are given on how schools were selected, contacts made, types of information collected, and techniques employed in analyzing the findings. The major portion of the book is then devoted to reporting characteristics studied, instruments developed and tested in collecting data, types of analyses made, and findings in relation to each of the several aspects studied. Over-all criteria by which decisions were reached concerning selection of factors to be studied or phases of student achievement to be tested are not reported. Nevertheless, the scope of the project is commendable and most of the problems attacked are very real to all who are concerned with achieving an effective home economics program in the high school.

Care was taken to insure the representativeness of the participating schools in relation to Minnesota as a whole and to match reimbursed and non-reimbursed home economics programs. Factors studied for their influence on the effectiveness of the program include the characteristics of the schools, of the pupils and their homes, and of the home economics teachers; the nature, organization and placement of content in the curriculum; and adequacy of facilities for teaching. While many aspects of pupil behavior were appraised, some types of learning considered important outcomes of a home economics program were not tested.

Results of the study are given not only in terms of factors which appear to influence the effectiveness of the home economics program in these schools, but also improvements noted during the period and means whereby teachers

and administrators had been challenged to bring them about. One of the strengths of the study is its demonstration of the values to be derived from participation in such a project.

The generalizations and recommendations drawn from the study prove exceedingly thought provoking. While the findings apply to a specific group of schools, they deserve serious consideration and should serve to stimulate continued investigation and experimentation. Those interested in home economics education research will find many leads for further study. Because of the controversial nature of some of the problems attacked and implications drawn, many will see need for further investigation to support or challenge the general applicability of the findings. Administrators, supervisors and all of those directly interested in the development of a functional high school home economics program will value this study for having clarified some very basic difficulties and hurdles to be over-come and for having demonstrated some possible approaches to strengthening the home economics program.

Ohio State University

Dorothy D. Scott

B  
Cattell, Raymond B., Factor Analysis: An Introduction and Manual for the Psychologist and Social Scientist (New York: Harper and Brothers, 1952), pp. xiii + 462, \$6.00.

This volume presents a non-mathematical introduction to factor analysis with considerable emphasis on the place of factor analysis in scientific method. The application of factor techniques to research problems in areas other than ability and achievement testing (non-psychological areas) is discussed in more detail here than in most books on factor analysis. Two of the main sections of the book are devoted to an elaborate and thought-provoking discussion of concepts and principles relating to factor analysis. The third is an extensive section on specific techniques for factoring and rotating. In addition, an appendix contains an outline of steps in matrix multiplication by electronic digital calculators.

The larger contribution of the book is probably in the ideas presented rather than in the working techniques. The "covariation chart", adding the time (or occasions) dimension to those of persons and variables illustrates the extension of factor analysis from R to what are labeled O, P, Q, S, and T techniques. Q-technique is regrettably treated with less sympathy than the other variations of R with little attention paid to its possible uses. The development of the concept of the "specification equation", in predicting behavior of individuals and groups in various situations is generally useful, being presented in more detail than the usual discussion of factor scores. Chapters coming late in the book on manipulation in the "classical factor analytic experiment" and on factor analysis with controlled experiment present material which should stimulate thinking among users of factor techniques.

The main criticism of the book stems from its planned appeal to the read-

er with little mathematical background. No attempt is made to tie in factor analysis with mathematical theory in general; in fact, mathematicians are chided for insisting upon some of the "niceties" in factor solutions. This is especially true in the presentation of the centroid method where use of some of the procedures outlined will not give a set of factors which is a true centroid solution. For example, the author suggests in his discussion of reflection procedures that it is not always desirable to maximize the sum of the matrix when the first centroid factors are being extracted. It is hoped that the casual reader will not attempt to follow the centroid method outlined in an early chapter not realizing that a more adequate presentation is made later in the book.

Group methods of factoring are treated in some detail but several available computational checks are not mentioned. Detailed numerical examples would have been a useful part of the outline of steps for diagonal factoring and for computing inverses. In the discussion of rotation procedures, the author treats the extended-vector approach very briefly suggesting that it is not as practical as radial rotation techniques. Existing techniques for direct rotation to primary structure are ignored completely. His dismissal of positive manifold (a criterion in rotation) as purely a myth implies some misunderstanding, i. e., that the criterion should be applied to all solutions, indiscriminately, rather than to solutions for matrices that are essentially positive at the outset.

Persons already familiar with factor analysis will probably object to the notation used, particularly V for variable and F for primary factor, since V and F have already taken on other meanings in the Thurstone school of factor analysis. Also, use of the terms "projection" and "loading" may be criticized although a proper distinction is made between "loading" and "correlation".

While not winning friends among psychologists who already reject factor analysis as lacking mathematical respectability, Cattell's Factor Analysis should find an enthusiastic audience among readers looking for ideas about the technique and among novices who are frightened by equations.

Combat Crew Training Research Laboratory  
Randolph Air Force Base, Texas

Dorothy M. Knoell

#### Books Received Recently\*

- Althouse, J. G. Structure and Aims of Canadian Education (Toronto: W. J. Gage and Co. Ltd., 1949), pp. 77, \$1.25.  
Baker, Harry J. Introduction to Exceptional Children (New York: The Macmillan Co., 1953), pp. xvi + 500, \$5.00.  
Butts, R. Freeman; Cremin, Lawrence A. A History of Education in American Culture (New York: Henry Holt and Co., 1953), pp. xi + 628, \$5.75.  
Faculties: Andover, Exeter, Lawrenceville, Harvard, Princeton, and Yale. General Education in School and College (Cambridge: Harvard University Press, 1952), pp. 142, \$2.00.

- Knapp, Robert H.: Greenbaum, Joseph J. The Younger American Scholar: His Collegiate Origins (Chicago: University of Chicago Press, 1953), pp. xiii + 122, \$3.00.
- Moustakas, Clark E. Children in Play Therapy: A Key to Understanding Normal and Disturbed Emotions (New York: McGraw-Hill Book Co., 1953), pp. ix + 218, \$3.50.
- Percival, W. P. Should We All Think Alike? (Toronto: W. J. Gage and Co. Ltd., 1951), pp. 112, \$1.25.
- Robbins, Florence Greenhoe. Educational Sociology (New York: Henry Holt and Co., 1953), pp. xiv + 529, \$4.75.
- Wiles, Kimball. Teaching for Better Schools (New York: Prentice-Hall, Inc., 1952), pp. xiii + 397, \$4.00.
- Ziegfield, Ernest. Art in the College Program of General Education (New York: Columbia University, Bureau of Publication, 1953), pp. x + 239, \$4.00.

\* Listing of books here does not preclude subsequent review.

# JOURNAL of EDUCATIONAL RESEARCH

Volume XLVII

November, 1953

Number 3

## A COMPARATIVE STUDY OF THE RELATIONSHIP BETWEEN THE QUALITY OF THE CHILD'S LANGUAGE USAGE AND THE QUALITY AND TYPES OF LANGUAGE USED IN THE HOME\*

DORIS IONE NOEL  
Anderson Public Schools  
Anderson, Indiana

MANY WRITERS on the subject of language have stressed the fact that language is largely a product of imitative behavior and that the environment, not the school, is the great determiner of the quality of language which the child is going to use. Since practically no research has been conducted to determine objectively the extent to which this might be true, this study was concerned with the development and use of some objective measuring devices for determining the relationship between the child's language usage and the usage which he heard in his environment. As the children were in the elementary grades, the parents were considered the most dominant environmental influence for children of this age. Thus, this study was designed to measure the relationship between the language usage of the child and the usage of the parents.

### Problem

In determining this relationship, the problem became a three-fold one: (1) to determine the relationship between the child's language usage and the quality of the language usage spoken by the parents in the presence of the child; (2) to determine the relationship between the language usage of the child and the occupation of the father; and (3) to determine the relationship between the quality of the child's language usage and the frequency with which his parents used oral expression.

### Method

With the problem defined, the next step involved the selection of

\*Unpublished Master's Thesis, Ball State Teachers College, 1951.

measuring devices which could be used in determining these various relationships. The writer's adviser, Dr. Robert H. Koenker,<sup>1</sup> who gave many valuable suggestions during the course of this study, suggested that the most objective of these devices would be: (1) an English usage test which could be given to the child and his parents; (2) an occupational classification of the father; and (3) a questionnaire which covered the various types of oral language situations.

Since none of the available standardized English tests contained an adequate sampling of the various usages, one had to be constructed. This was done following the analysis of the usage items which were covered by Charters,<sup>2</sup> DeBusk,<sup>3</sup> the language guide for the writer's school system,<sup>4</sup> most of the standardized English Tests, and the language tests included in most of the standardized achievement batteries.

The fifty usage items which were common to four or more of these sources were the basic items upon which the test was constructed. The final test contained two hundred pairs of sentences. One of the pair was correctly stated and the other was incorrectly stated. Both forms were used in order to preserve any feeling of familiarity with either expression in the belief that a person might more nearly check the expression he used if he did not have to stop to think before he marked the correct form. The test covered the following eight categories of usage: "verb agreement", "verb tense", "wrong tense form", "verb often confused", "pronouns", "adjectives or adverbs", "redundancy" and "prepositions".

The second measuring device, the occupational classifications of the fathers, was based upon the classifications as listed in the Dictionary of Occupational Titles.<sup>5</sup> The classifications for this study were

<sup>1</sup> Director of Graduate Studies and Associate Professor of Education, Ball State Teachers College, Muncie, Indiana.

<sup>2</sup> W. W. Charters, "Minimal Essentials in Elementary Language and Grammar," Second Report of the Committee on Minimal Essentials in Elementary School Subjects, pp. 85-110. The Sixteenth Yearbook, Part I. National Society for the Study of Education. Bloomington, Ill.: Public School Publishing Co., 1923. 204 pp.

<sup>3</sup> Burchard Woodson DeBusk, The Persistency of Language Errors Among School Children, University of Oregon Bulletin Vol. 2, No. 4. Eugene, Oregon: University Press, May, 1930.

<sup>4</sup> Suggested Guide for a Language Program in Anderson Public Schools, 1950.

<sup>5</sup> Part I, Definition of Titles, 1287 pp.; Part II, Group Arrangement of Occupational Titles and Codes, Washington: United States Government Printing Office, June, 1939. 330 pp.

not so specific, however, since the number of cases was limited when the children were also matched on intelligence quotients.<sup>6</sup> The four occupational groups which were used were: (1) Professional or Managerial; (2) Clerical, Sales, or Service; (3) Skilled; and (4) Semi-skilled or Unskilled.

The third measuring device, the questionnaire, was used as a guide in interviewing parents to ascertain the extent to which they participated in situations requiring the use of oral expression. It was constructed to determine whether parents "always", "often", "seldom", or "never" took part in the activities listed under the general headings: "Conversation", "Use of the Telephone", "Introductions", and "Participation in Clubs or Other Organizations". A point value, ranging from three points for answers of "always" to no points for answers of "never" was arbitrarily assigned to each answer. The sum of these points was then used as the measure of parental participation in oral language situations.

Since the pupils who were taught by the writer lived in a highly industrialized section of Anderson and the parents represented all but the wealthy economic level, it was decided that they would be good subjects for this study. Therefore, at the suggestion of her former supervisor, Miss Grace A. Klein, and Superintendent A. R. Chadd, the writer appeared before the Anderson Board of School Trustees, outlined the study, and received permission to seek the cooperation of the parents in making the investigation.

Letters were sent to 155 homes from which came the 177 children in the writer's departmental English classes, grades four A to six A inclusive. These letters explained the purpose of the study and asked the parents to indicate, on the form which was enclosed, whether they would be willing to have the writer visit with them in their home and give them the English Usage Test. Replies were received from 86% of the families and all but ten indicated their willingness to participate in the study.

Visitations were begun on December 1, 1950 and were terminated at the close of the school term on June 8, 1951. During this time 107 families which were represented by 124 pupils, or 71% of all pupils in the departmental classes, were visited. (Illness in the writer's own family prevented her from visiting all the parents who wished to cooperate.) No school time was used in making these visitations which averaged one hour and forty-five minutes. The parents were very cooperative in answering the points on the questionnaire and in marking the sentences which they said in the presence of their children rather

<sup>6</sup>Elizabeth T. Sullivan, Willis W. Clark, and Ernest W. Tiegs, California Short-Form Test of Mental Maturity, Elementary, Grades 4-5-6-7-8, 1950 S-Form, Los Angeles: California Test Bureau, 1950.

than the ones which they knew to be correct.

The children were given the test prior to the visitations so that the reliability and validity of the test were established. The reliability coefficient<sup>7</sup>, determined by the split-half technique, was .96 which indicates that the English Usage Test was a reliable measure of language usage. The validity coefficient<sup>8</sup>, established by correlating children's scores on the writer's test against their scores on the Clapp-Young English Test - Form A<sup>9</sup> was .85. This indicates that the test used in this study was a valid measure of language usage.

### Findings

After all the tests for the children and parents had been scored, the questionnaires tabulated, the occupational classification determined for each father, and the intelligence quotient figured for each child, the data was treated statistically by means of Pearson's Product Moment Coefficient of Correlation and the standard error of correlation<sup>10</sup>, "t" tests for the significance of the difference between the means<sup>11</sup>, and percentage comparisons.

Most of the findings were based upon the errors made by fifty children, selected at random, and their mothers and fathers, since it was found, by comparing the coefficients of correlation<sup>12</sup> on the total test for all the cases (.46) and the sampling (.49), that the sampling was representative of the entire group.

Correlations<sup>13</sup> and "t" tests<sup>14</sup> were run on the number of errors made by the child and each of his parents on the total test of two hundred sentences and each of the eight types of usage covered by the English Usage Test. Correlations<sup>15</sup> were also made between the child's score

<sup>7</sup>Charles C. Peters and Walter R. VanVoorhis, Statistical Procedures and Their Mathematical Bases, pp. 92-98. New York: McGraw-Hill Book Co., 1940. 363 pp.

<sup>8</sup>Ibid.

<sup>9</sup>Boston: Houghton Mifflin Co., 1929.

<sup>10</sup>Peters and VanVoorhis, op. cit., pp. 92-98.

<sup>11</sup>E. F. Lindquist, Statistical Analysis in Educational Research, New York: Houghton Mifflin Co., 1940. 257 pp.

<sup>12</sup>Peters and VanVoorhis, op. cit. pp. 92-98.

<sup>13</sup>Ibid.

<sup>14</sup>Lindquist, op. cit., pp. 37-40.

<sup>15</sup>Peters and VanVoorhis, op. cit., pp. 92-98.

on the English Usage Test and his parents' score on the questionnaire. The "t" test<sup>16</sup> was applied to the difference between the means of the English Usage Test scores for the children according to the four occupational classifications. Percentage comparisons were made to determine the usages which were the most difficult and to determine the extent to which the errors made by the child were also made by one or both parents.

The most significant findings which resulted from these comparison were:

1. Significant product moment coefficients of correlations<sup>17</sup> of .45 and .52 showed that there was a direct, positive relationship between the number of errors made by the child and each of his parents on the total test. The "t" values<sup>18</sup> for these same comparisons were less than the .01 level of probability<sup>19</sup>. These indicate, as do the coefficients of correlation, that there is a definite relationship between the quality of language usage spoken by the parents and the quality of the language usage of the child.

2. A further breakdown of the total number of errors into the number of errors which were made under the eight different types of usage showed that there were no significant differences between the number of errors made by the child and either parent on the sentences testing "verb tense", "wrong tense form", "pronouns", "adjectives or adverbs", and "redundancy". This also indicates, in general, that the child tends to use the same quality of language usage as his parents do.

3. Only in sentences testing the use of the "prepositions" was there a significant between the number of errors made by the mother and the child. Only in sentences testing the use of "prepositions", "verb agreement", and "verbs often confused" were there significant differences between the number of errors made by the father and the child. This indicates that in the use of prepositions, verbs to agree with the subject in number, and verbs which are often confused, there is little relationship between the English usage of the child and that of his parents.

4. Therefore, in five of the eight types of language usage, the children tend to use the same quality of English as that which they hear their parents use.

<sup>16</sup> Lindquist, op. cit., pp. 37-40.

<sup>17</sup> Charles C. Peters and Walter R. VanVoorhis, Statistical Procedures and Their Mathematical Bases, pp. 92-98.

<sup>18</sup> E. F. Lindquist, Statistical Analysis in Educational Research, pp. 37-40.

<sup>19</sup> Palmer O. Johnson, Statistical Methods in Research, New York: Prentice-Hall, Inc. 1950. 377 pp.

5. Children made a slightly lower percentage of errors than did their parents in the usages involving "verb agreement", "wrong tense form", "verbs often confused", "adjectives or adverbs", and "prepositions". They made a slightly higher percentage of errors than their parents in the use of "verb tense" and "pronouns". They also made a slightly higher percentage of errors than their mothers but the same percentage of errors as their fathers in the use of redundant expressions.

6. Errors were fairly well spread over the entire eight usages; however, individual items in the use of the verb and the pronoun accounted for the ten most frequent errors made by the children. Seven of these ten errors were also found to be among the ten most frequent errors for one or both parents. These errors were made in the use of "ring", "drownded", "rise-raise", pronouns as subjects, predicate pronouns, pronouns as objects of prepositions, and "who-whom".

7. Applying this arbitrary standard, whenever more than 75% of the errors which were made by the child were also made by either one of the parents, the language habits of the child were considered the result of home environment; it was found that the language habits of 64% of the children in the sampling were determined by the home.

8. The "t" tests<sup>20</sup> on the differences between the means of the English Usage Test scores for the children grouped according to the four occupational levels of the fathers showed that there were no significant differences between the means of any of the occupational groups. In general, the occupation of the father did not affect the quality of the language usage of the child.

9. From the comparison of the English Usage Test score of the child and his parents' score on the questionnaire, it was found that there was a direct, positive relationship between the quality of language used by the child and the frequency with which his parents used the different types of oral expression. This further indicates that the quality of the language which is used by the parents has direct bearing upon the quality of language used by their children.

### Conclusions

The major conclusions which can be drawn as a result of this study are:

1. The language usage which the child in the elementary grades hears his parents use does, to a very large degree, determine the quality of language usage which the child uses.

2. The teacher can, by constant practice on certain types of usage,

<sup>20</sup> Lindquist, *op. cit.*, pp. 37-40.

help the child eliminate some of his errors in these usages.

3. The parents must cooperate with the schools in seeing that the children hear and practice correct English in the home if much improvement is to be made in the quality of language used by their children.

4. When the intelligence quotient factor is held constant, the occupation of the father does not materially affect the quality of language used by the child.

5. In general, the more frequently the parents participate in situations requiring the use of oral expression, the better will be the quality of the child's language usage.

#### Limitations of the Study

The above conclusions are subject to the following limitations:

1. The degree to which any language test can be said to measure a person's control of the English language is subject to question.

2. Evidence that the parents checked the test sentences the way they said them rather than the way they thought the writer might want them to check the sentences can not be determined scientifically.

3. These factors, which may have an effect upon the language usage which the child uses were not controlled:

- a. The amount of time spent with the parents and the amount of time spent outside the home environment
- b. The presence or absence of siblings
- c. The type of language used by the close associates of the child
- d. The difference in the amount of control children on different grade levels have over language usage
- e. The rather broad occupation classifications used for this study

# ACHIEVEMENT, APTITUDE, AND BACKGROUND OF LIBERAL ARTS AND SCIENCE STUDENTS (GENERAL CURRICULUM) DEFICIENT IN HIGH SCHOOL MATHEMATICS

CLARENCE PHILLIPS

University of Illinois

Urbana, Illinois

THE QUESTION of preparation for college is a matter of mutual concern of colleges and high schools. Especially is this true for the students who enter the liberal arts and science colleges. In order to set up minimum basic knowledge requirements, colleges may have entrance requirements and/or graduation requirements.

University of Illinois Liberal Arts and Science freshmen in the general curriculum lacking one year high school algebra and one year high school plane geometry are termed "deficient in high school mathematics". In order to graduate in the general L. A. S. curriculum, these students are required to take a two semester basic mathematics course earning a total of 6 semester hours. In 1949-50, 113 students entered the University of Illinois L. A. S. general curriculum deficient in high school mathematics. In 1950-51, the number was 121. This represents approximately 10% of each year's entering class in the Liberal Arts and Science College at the University of Illinois. Of this 10% deficient in high school mathematics, 60% are women. In reference to the entire L. A. S. freshmen class 1950-51, 50% were women.

In order to determine the content of this basic mathematics course for the students deficient in high school mathematics, it is essential to know the achievement, aptitude, and background of these students. This article will present data on these students (1949-51) as they enter the University of Illinois L. A. S. general curriculum. The following topics will be considered:

1. Mathematics achievement
2. Achievement and aptitude in reference to the entire L. A. S. freshmen class
3. Scholastic record in high school
4. Personal background
5. Correlations

The following tests were given to ascertain achievement and aptitude:

1. Hundred-Problem Arithmetic Test by Schorling and Clark.

TABLE I

## ITEM ANALYSIS OF MATHEMATICAL KNOWLEDGE

1.	Whole numbers	
a.	Column addition, 7 numbers ranging from 2 to 5 digits	77%
b.	Division of a 3 digit number by a 2 digit number. No remainder nor zero involved	77%
2.	Fractions	
a.	Add 3 proper fractions involving fourths to sixteenths	60%
b.	Subtract 2 mixed numbers involving "borrowing"	46%
c.	Multiply mixed number by a whole number	57%
d.	Divide a fraction by a whole number	46%
3.	Decimals	
a.	Divide 4 digit number (involving tenths) by a 2 digit number (involving hundredths) and no remainder	55%
b.	Arranging in size (hundredths, mixed number, thousandths)	30%
4.	Measurement	
a.	Find volume of a rectangular solid with 3 dimensions given in feet (integers)	63%
b.	Find circumference of a circle with diameter given (integer)	26%
c.	Find volume of a cylinder with the formula given and the dimensions in feet	32%
5.	Number system	
a.	Recognizing a given number as sum of integer, tenths, and thousandths	54%
b.	Rounding off ten-thousandths to the nearest hundredth	50%
6.	Word Problems	
a.	Finding the cost of 4 articles from the cost of a dozen given	53%
b.	Simple interest for a year involving integers	41%
c.	Simple interest for a year involving mixed number as a per cent	24%
d.	Retail discount involving the use of the following: original price minus discount taken off the original price (all integers)	38%
e.	Finding the rate of increase with the principal and increase given	19%
7.	Geometry	
a.	Using the knowledge that the sum of the angles of a triangle equal 180 degrees	50%

- World Book Co., 1942.
2. American Council on Education. Cooperative General Achievement Tests. Mathematics, Natural Science, Social Science, and Reading. Cooperative Test Service, 1947.
  3. American Council on Education. Psychological Examination for College Freshmen. American Council on Education, 1946. Used by the University of Illinois guidance bureau as a test of scholastic aptitude.

#### A. Mathematical achievement

##### 1. Arithmetic

The median for the L. A. S. freshmen (general curriculum) deficient in high school mathematics is 57% on the Hundred-Problem Arithmetic test. According to the authors of the test, the median for students in the eighth grade is 62% (manual of directions p. 6).

##### 2. General achievement in mathematics

For the general achievement in mathematics, the median for this group of students deficient in high school mathematics is 29% on part I, and 36% on part II on the Cooperative general mathematics achievement test.

For the general achievement in mathematics on the Cooperative test, the median for this group of students deficient in high school mathematics is a scaled score of 41. The median for college freshmen entering liberal arts colleges according to the publishers of the test is a scaled score of 58.2.<sup>1</sup>

#### B. Item analysis mathematical achievement

In order to ascertain the level of mathematical achievement on various topics, an item analysis was made for the various answers given on the 3 tests listed above. For this article only certain problems given at the beginning of timed tests are used. In this manner everyone had an opportunity to try to solve the problem.

This item analysis shows that only 46% of the L. A. S. freshmen (general curriculum) deficient in high school mathematics are able to divide a proper fraction by a whole number. In Table I the percent given indicates the percent of L. A. S. freshmen (general curriculum)

<sup>1</sup>Report on the 1950 National College Freshmen Testing Program. Cooperative Test Division. Educational Testing Service.

TABLE II

**ACHIEVEMENT AND APTITUDE RANK IN REFERENCE TO ENTIRE L. A. S.  
FRESHMEN CLASS UNIVERSITY OF ILLINOIS**

	Per cent in			
	4th quartile	3rd quartile	2nd quartile	1st quartile
<b>Cooperative Achievement</b>				
Mathematics	2	7	27	64
Natural Science	7	12	29	52
Social Science	11	14	29	46
Reading Level	10	21	35	34
Scholastic aptitude	8	8	29	55

TABLE III  
ACHIEVEMENT AND APTITUDE  
IN REFERENCE TO NATIONAL FRESHMEN TESTING PROGRAM

	L. A. S. Freshmen Deficient in High School Math.	National 1950 Freshmen
Natural Science	56 scaled score	60.5 scaled score
Social Studies	57 " "	61.0 " "
Reading Level	53 " "	55.6 " "
A. C. E. Psych. (Total)	91 raw score	107.3 raw score

TABLE IV  
GRADE AVERAGE AND PERCENTILE RANK IN HIGH SCHOOL

---

High school grade average for 4 years:

A 7%	B 44%	C 46%	D 3%
---------	----------	----------	---------

Percentile rank high school graduating class:

4th quartile 30%	3rd quartile 42%	2nd quartile 20%	1st quartile (0-25) 8%
---------------------	---------------------	---------------------	------------------------------

---

deficient in high school mathematics who solved the problem correctly.

**C. Scholastic aptitude and achievement in reference to the entire L. A. S. freshmen class**

Considering mathematical achievement, 64% of the L. A. S. freshmen (general curriculum) deficient in high school mathematics are in the first quartile (0-25 percentile) of the entire L. A. S. freshmen class. Note however that 9% rank in the upper half of the entire L. A. S. freshmen class in reference to mathematical achievement. Table II shows the placement of this group in reference to the entire L. A. S. freshmen class.

In reference to achievement and aptitude the group of students in L. A. S. (general curriculum) deficient in high school mathematics ranks slightly lower than the college freshmen tested on a national wide basis.<sup>2</sup> Table III gives the medians for the two groups.

**D. Scholastic record in high school**

Considering the four year high school record of the L. A. S. freshmen (general curriculum) deficient in high school mathematics, 49% had an average less than B. In contrast 28% ranked in the lower half of their high school graduating class. Table IV gives this information.

Sixty-four percent of the L. A. S. freshmen (general curriculum) deficient in high school mathematics have the required one year<sup>2</sup> high school algebra credit. However only 14% have had 1/2 to 1 unit of high school plane geometry. See Table V.

Sixty-two percent of the L. A. S. freshmen (general curriculum) deficient in high school mathematics had less than a B average in algebra or geometry or both. Note if the student had both algebra and geometry, then the latter was only a half unit.

A	B	C	D	Fail
8%	30%	37%	24%	1%

**E. Personal background**

Table VI shows the percent of students coming from various sized communities. For example 24% of the L. A. S. freshmen (general curriculum) deficient in high school mathematics come from cities of 1 million or over. In comparison, 33% of the entire L. A. S. freshmen class come from cities of 1 million or over.

Thirty-two percent of the L. A. S. freshmen (general curriculum)

<sup>2</sup>Ibid.

TABLE V  
HIGH SCHOOL MATHEMATICS SUBJECTS

---

Advanced Math. (Algebra, Trig., 1/2 to 1 unit)	8%
Algebra. (1 unit)	64%
Geometry (1/2 to 1 unit)	14%
General Math. 9th grade	23%
Practical Math. (1/2 to 1 unit)	8%
Commercial Arith. Business Training (1/2 to 1 unit)	21%

---

TABLE VI  
POPULATION OF L. A. S. FRESHMEN HOME COMMUNITY

Population	Entire Class 1950-1	Group Deficient in H. S. Math. (1949-51)
1,000,000 and over	33%	24%
Suburbs of Cities		
1,000,000 and over	9	7
100,000 - 900,000	4	9
50,000 - 90,000	5	6
25,000 - 49,000	6	9
15,000 - 24,000	7	11
10,000 - 14,000	6	5
5,000 - 9,000	6	7
1,000 - 4,000	9	8
Below 1,000	8	6
Rural	7	8

TABLE VII  
OCCUPATION OF FATHER\*

---

Professional and Managerial	32%
Clerical and Sales	12
Service	8
Agriculture	8
Skilled	14
Semi-skilled	16
Unskilled	10

---

\* Classifications used: U. S. Dictionary of Occupational Titles. Washington, D. C. Bureau of Employment Service, 1949. v. 2, ix.

TABLE VIII  
CORRELATIONS

---

1. Achievement in Mathematics		
Hundred-Problem Arith. Test		
" " " "		
A. C. E. Psych. Test		
2. Achievement Tests (Cooperative)		
Mathematics		
"		
"		
3. Word Problems		
Word Problems (A. C. E. Psych. Test)		
4. High School Record		
High school math. average		
(Algebra and Geometry)		
Percentile rank graduating class		
	-Coop. Math. Achievement	.47
	-High school math. average (Algebra and Geometry)	.39
	-Coop. Math Achievement	.59
	-Natural Science	.61
	-Social Science	.54
	-Reading Level	.40
	-Reading Level (Coop.)	.46
	-High school 4 year average	.74
	-High school 4 year average	.79

---

deficient in high school mathematics come from homes where the father's occupation is classed as Profession or Managerial. See Table VII.

#### F. Correlations

Various correlations were computed in order to find out the relationship between the various tests and high school record. All of these correlations are in reference to the L. A. S. freshmen (general curriculum) deficient in high school mathematics. See Table VIII.

In reference to the above data for the University of Illinois L. A. S. freshmen (general curriculum) deficient in high school mathematics, the following observations may be stated:

1. Most of the students are extremely low in mathematical achievement. This even extends into arithmetic. However approximately 10% rank in the upper half of the entire L. A. S. freshmen class.
2. Most of the students are low in all subject matter achievement. Here again we should note that exceptions occur.
3. Most of the students are low in scholastic aptitude. However approximately 15% rank in the upper half of the entering L. A. S. freshmen class.
4. The students do not come from a specific sized community. Instead the range is distributed from rural to cities over a million. The occupation of the father is also distributed from the unskilled to the professional and managerial class.
5. There is a low correlation between student's high school record and mathematical achievement.

# VOCABULARY TESTS AND DEPTH OF MEANING

E. W. DOLCH  
DON LEEDS  
University of Illinois

HOW WELL do our existing vocabulary tests measure children's knowledge of word meaning? That was the question which started a study which has led to interesting results and conclusions, and that has raised still more questions about the measuring of children's word knowledge.

The first approach was to compare existing vocabulary tests. Five tests were studied, three by authorities in the field of reading, and two which are in nationally used batteries of achievement tests. The tests were the Thorndike, Gates, Durrell-Sullivan, Stanford, and Metropolitan. The words used in these tests were listed by parts of speech and close agreement was found as to the percentages of each of the parts of speech included.

A sampling was then made of a widely-used school dictionary, a sampling that equalled the total number of words used on all the vocabulary tests studied. A school dictionary was used instead of the large unabridged dictionary because, after all, vocabulary tests are intended for school and there should be some relation therefore between vocabulary tests and a dictionary used by children. The words of the sampling were then listed by parts of speech using as the part of speech for each word the one given first by the dictionary, and the percentages compared with the percentages of the total of the vocabulary test list. Results appear on Table 1.

The table shows not too great a difference between the vocabulary test list and the dictionary sampling. The greater weight given by the tests to verbs might have been caused by the fact that very often a word may be used both as a noun and as a verb, and the tests might have given preference at times to the verb meaning because it was easier to make a test item for it. For instance on one vocabulary test appeared the words feast and annex. At first glance these would seem to be nouns but it would be hard to make a short definition of them as nouns. Instead the test defined them as verbs and used as synonyms the verbs "to eat" and "to add". At any rate, the vocabulary tests agree rather well with a sampling of a single dictionary.

TABLE I

**COMPARISON OF PERCENTAGES OF PARTS OF SPEECH ON  
DICTIONARY SAMPLING AND VOCABULARY TEST LIST**

	N	V	Adj.	Adv.	Prep.	Interj.
Dictionary	58.83	15.08	21.98	3.87	.21	....
Vocabulary Tests	45.64	29.52	21.72	3.03	.002	.001

Different-sized dictionaries, made according to different policies, might have given different results.

But the dictionary gives many meanings of words, and not just one meaning as given by a vocabulary test. How were the tests related to the dictionary on this point? Each one of the 770 words contained in all the vocabulary tests was looked up in the dictionary. It was found that, in general, the vocabulary tests used as the "right" answer, the first definition given in the dictionary. This particular dictionary follows the practice of giving first what the makers considered to be the "most common meaning" of the word, in contrast with some dictionaries which give first the original or historically first meaning of the word. Our conclusion must be, therefore, that in general the standard vocabulary tests are measuring a sampling of the school dictionary, especially the most common meaning of each word tested.

What about the further meanings given by the dictionary? Apparently the standard vocabulary test can help us very little with them. These further meanings are of at least three kinds.

1. Same Spellings from different roots. In the dictionary we find many cases where the same spellings come from different roots and therefore have quite different meanings. For instance, there are three different words all spelled "box". One word means a container, the second word means to strike with the open hand, and the third word means a small tree used in gardens. The dictionary shows these as separate entries, with small serial numbers beside them. Of the dictionary sampling we took, 6.28% of the words had other words of this type, with the same spelling but with different meaning. Naturally, these "same spelling" words cannot be fully tested by the usual vocabulary test. The school considers them in the language course under the study of homonyms.

2. Derived meanings. Many words in the dictionary have, in addition to the most common meaning, other meanings which are derived from it through analogy or through application to various special fields of use. In the case of one of the words spelled "box", one meaning is "the amount a box can hold", as "he can eat a whole box of candy". This meaning is derived from the size of the box or container. The meaning "to box an article for shipment" comes from the action connected with the idea of a box. The meanings "a driver's seat on a coach, a theatre box, a sentry box" are derived from a similarity to the form of a box. There are endless methods of deriving meanings from the fundamental or basic or first meaning. Vocabulary tests have so far not entered this field of derived meanings.

3. Figurative meanings. There are endless figurative meanings also for most words. We often say, "I was certainly in a box" when

we really mean "predicament" but use the figure of "box" instead. Some of these meanings are given in the dictionary and some are not. The larger the dictionary, the more of them are given. But no dictionary can include all the possible figurative meanings that writers may think of for words. Naturally, the vocabulary test does not include figurative meanings.

To summarize, we found that the vocabulary tests almost universally used only the first or most common definition of a word, leaving out, 1. homonyms, 2. derived meanings, and 3. figurative meanings.

But did the standard vocabulary tests actually measure this first or most common meaning? The next step in the study was to try to answer this question.

### Depth or Degree of Meaning

The homonyms, derived meanings, and figurative meanings just mentioned may be thought of as degrees of meaning beyond the most common meaning. But the most common meaning has also degrees of meaning, or different depths of meaning. Our question is, how well do vocabulary tests measure depth in the case of this most common meaning? To try to answer this question, an intensive study was made of the way one of the vocabulary tests tried to test word meaning. This test was chosen because of its wide use and general excellence.

This test used two devices to measure word meaning: (1) synonyms and (2) the classification or part-whole method. Let us look first at the classification method as it was used with nouns. The meaning of 18 nouns was measured as follows:

### Meaning of Nouns by the Classification Method

Test Word	Tested by Marking the Class	Test Word	Tested by Marking the Class
red	color	dome	roof
rabbit	animal	portal	door
mother	woman	mane	hair
cow	animal	pew	seat
orange	fruit	dahlia	flower
shower	rain	desertion	leaving a position
pole	stick	spinster	woman
lane	path	insignia	badge
colonel	officer		
murderer	killer		

(Two nouns were tested by this method in reverse. That is, the words were vehicle and projectile, which are classes of things, and they were tested by marking one thing in each class, i.e. carriage and bullet)

In the case of these nouns, for one to know the class to which each belongs may be a first degree or minimum meaning. It would help in reading, especially if the context aided. But one who knew only this first degree of meaning would have a very vague idea of the word and could not use it very well in either speaking or thinking. Suppose all one knows about an orange is that it is "some kind of a fruit" or all the meaning one has for colonel is "some kind of officer". He could not think intelligently about either. So this method of testing the minimum meaning is useful to some degree for reading but not for other language use.

#### Depth of Meaning by the Synonym Method

By far the commonest method of testing word meaning on this standard test was by synonyms. In fact synonyms were used 76% of the time in the entire test. The problem of testing meaning by synonyms is well known in the definition of synonyms given in the New International Dictionary as follows: "Synonyms are words which express what is essentially the same idea but which (commonly) differ from one another, (1) in some shade of meaning, (2) in emphasis or (esp.) (3) in their connotations." The synonyms given by the test for 14 nouns are below:

#### Meaning of Nouns by the Synonym Method

<u>Test Word</u>	<u>Synonym given as the meaning</u>
disaster	ruin
gap	opening
splendor	glory
bravery	courage
strife	fighting
comment	remark
flaw	fault
malady	sickness
symptom	sign
avalanche	landslide
carousal	revel
lethargy	apathy
animation	liveliness
bouyancy	lightness

Does the checking of a synonym tell that one knows the meaning of a word? First, we have to assume that the testee knows the meaning of the synonym and has not just heard the two words together so that he merely associates them in some way. If the testee does know the meaning of the synonym, he at least knows the degree or depth of meaning carried by that synonym. To know what that meaning is, would require still further testing. If the meaning the testee has for the synonym is slight the meaning of the word is just as slight.

Second, does the testee know the difference between the meaning of the synonym and the meaning of the test word? The vocabulary test obviously cannot answer that question.

For the other parts of speech on the test, the method was in all cases synonyms. As the problems with them are the same as with the problem of synonyms of nouns we shall not present the results here because of lack of space.

### Testing of Further Depth of Meaning

Having seen that vocabulary tests attempt to test merely the commonest meaning of a word, and test that either by giving a classification or a synonym, we were interested to know if students who had gone through our school system knew more than just this minimum meaning or synonym meaning of these words. Could we test further meaning of the words?

To attempt to test "further depth of meaning" we constructed a test based upon the standard vocabulary test just referred to. Our Depth of Meaning test gave the testee, in each instance, the information that the standard test required and then asked for more meaning.

For instance the following items are typical:

#### Classification Method

##### Standard Test

- A cow is
  - animal
  - shy
  - red
  - fish
  - coat

- A colonel is
  - officer
  - medicine
  - harness
  - poetry
  - disease

##### Depth of Meaning Test

- A cow is an animal that
  - a. is found in zoos
  - b. is used for racing
  - c. gives milk
  - d. does not have calves

- A colonel is an officer who commands
  - a. a battleship
  - b. a police department
  - c. a regiment of soldiers
  - d. a company of soldiers

Synonym Method

Disaster is	A disaster is ruin that happens
find	a. suddenly
shot	b. within a year's time
ruin	c. to all people
fence	d. gradually
object	
Lethargy is	Lethargy means apathy but
apathy	refers more to
wild animal	a. the body than the feelings
debt	b. men than to women
liveliness	c. sleep than to waking
	d. the soul than to the body

As will be noted from the examples, the Depth of Meaning Test in each case admitted possession of the knowledge asked for by the Standard vocabulary test, and then asked for more. In the case of classification it asked for some difference from others in the same classification. In the case of synonyms, it asked for the difference between the synonyms. The difference required as correct was that given in the New International Dictionary, both in the definitions of the two words and in the paragraphs explaining the differences between synonyms.

The Depth of Meaning Test was given to groups of college freshmen, college seniors and teachers in service. Since this study is merely exploratory, it will be sufficient to give the general results and we will do this only for the nouns, since the verbs and adjectives gave about the same result.

Results of Depth of Meaning Test

For most of the nouns which the standard test measured by classification, the adults had learned also the further meaning required by the depth of meaning test. On some of the words, however, they definitely had not. In this group, the percentages of adults knowing the further meaning required were:

shower	57%
lane	63%
colonel	70%
portal	47%
dahlia	27%

In the case of shower, many adults thought the word meant a light rain, whereas the dictionary very definitely says it means a short rain, whether light or heavy. For lane, we required the country meaning but many adults knew only the traffic lane meaning. For portal, the adults were very vague as to how it differed from door. And they just did not know what a dahlia was like.

In the case of the synonyms it seemed definitely hard for the adults to differentiate between them. Of the ones we have listed, four were easy: disaster, symptom, avalanche, and buoyancy. But about one out of four adults could not distinguish between the synonyms for gap, splendor, bravery, comment, and flaw. And about half were wrong when trying to distinguish between synonyms for strife, malady, carousal, lethargy, and animation.

The same situation held true with the synonyms given by the standard test for verbs and adjectives. Space does not permit samples.

#### What Should We Expect of a Vocabulary Test

First of all, we must admit that existing vocabulary tests were not trying to measure scientifically a certain degree of meaning. They were just trying roughly to compare a child with other children of his own age with regard to the general field of vocabulary. We therefore are not criticising the vocabulary tests for doing something they did not try to do.

Having made this admission, we need to go on and to say that existing vocabulary tests certainly do not do a scientific job of measuring vocabulary. They ignore all but the most common meaning of a word, and then in cases of classification they test very little of that most common meaning, and when they use synonyms they test a very indefinite amount of knowledge.

It is to be hoped that in the future there will be an attempt made to measure scientifically what meaning children and adults have for words. Since words are the symbols with which we communicate with one another, it is important to know just what meaning or meanings words have for different people. Since words in reading are the symbols by which we try to get the thought a writer tried to put into a book, it is important that we have the right meaning and adequate meaning.

In this future scientific study, the first consideration will be that a word does not just have a "meaning". That naive assumption is all too prevalent even among educators. Instead, a word frequently has many meanings, and each of these meanings may be known in varying degrees by different people, or by the same person in the course of his life.

For each person, therefore, meaning is a growth, and it should be important for the school to know where any certain child is in that process of growth. How otherwise can the school wisely and efficiently help in the growth of meaning? Tests should tell us not merely that a child "knows something about" a certain word, but also how much he knows about it. Only then do we know what more he should know and how we can help him develop that added meaning.

# THE RELATIONSHIP BETWEEN INTELLIGENCE AS DETERMINED BY INTELLIGENCE TESTS AND THE ABILITY TO LEARN AS DETERMINED BY PERFORMANCE ON LEARNING TESTS

CLINTON WALLACE GREEN  
University of Pennsylvania

## Introduction (Need for Study)

INTELLIGENCE testing in some form is perhaps as old as human intelligence, but recent developments have made a creditable contribution in methods of measuring it. Since intelligence or native ability cannot be measured directly, test makers and persons interested in measuring it have sought to quantify it in terms of its manifestations and symptoms. For the most part, instruments devised for that purpose are based upon the assumptions that all individuals within a given age limit have had comparable experiences and that individuals acquire those habits, skills, knowledges, and abilities proportional to their native intelligence.

It was Alfred Binet of France, who first published an intelligence scale (1905) which he thought would measure the "higher mental processes" in an effort to select the mentally retarded children who required special instruction and who could not benefit by the usual instruction offered in the Parisian schools.

Many definitions of intelligence have appeared but it is generally defined and referred to as the ability or capacity to learn.

Upon examining several of the now popular and widely used intelligence tests, the writer concluded that one's score on these tests would tend to be influenced greatly by one's schooling, previous educational opportunities and cultural background. Intelligence tests are largely used to predict success in school. This tendency may be explained by noting that it is the schoolmen who construct these instruments, they are intended for persons of school age, and they are used most extensively in the schools. The fact that many instruments exist for school age groups and few are available for the measurement of adult intelligence indicates that the school age is foremost in the minds of the test makers. Wechsler discusses this point at length. (9)

Since intelligence tests purpose is to furnish information to supplement that secured from achievement tests, it seems logical to conclude that intelligence tests answer the question: How much can one learn? It is reasonable to assume that this question is foremost in the minds of

those who use intelligence tests.

The writer felt that much of the material now contained in intelligence tests was not equally familiar or equally new to all individuals and sought to measure performance with material when the initial point was the same for all of the subjects.

### Statement of the Problem

This study seeks to answer the question: What relationship, if any, exists between the ability to learn as determined by intelligence tests and school marks on the one hand, and certain learning exercises on the other.

### Related Research

Johnson gave 60 college students a 10 minute practice period every day for 20 days in mirror reading. During the course of the experiment several intelligence tests were given to the students. He obtained correlations between the average scores on all tests and performance in mirror reading of  $.34 + .08$  for the average number of words read each day, and  $.46 + .07$  with improvement in mirror reading over 20 days. He found that students in the upper half on intelligence improved more in mirror reading than those in the lower half. He concluded that there was a fairly large positive relation between the ability to become efficient at learning inverted print and intelligence as measured by the usual group tests. (4)

Peterson found a moderate correlation between his rational or mental maze test and scores on intelligence tests. He gave no statistics to show the relationship but stated that the bright pupils tended to respond in a definite manner. (7)

Jordan administered four group intelligence tests (Otis, Alpha, Miller, and Terman Tests), the Stanford Revision of the Binet Scale, and his educational learning test to 64 high school pupils. The correlations between intelligence test scores and the learning test scores were uniformly low for the group. They centered around .20, and ranged from  $.31$  to  $-.125$ . (5)

### A Description of the Tests

The learning test employed in this study consists of five parts, Letter Observation, Digit-Symbol, Vowel-Consonant, Parentheses Marking, and Reversed Type. These tests were selected and/or designed to give pupils an opportunity to demonstrate their ability to learn. It is believed that previous schooling, and acquired information would operate only indirectly to influence one's score. Each exercise was further

designed or selected so that practice would make a difference, *i. e.*, one would tend to find the exercise easier and easier as one goes on.

The Letter Observation Test is based upon the principle embodied in the Kellogg-Morton Revised Beta Examination and the Pintner-Cunningham Primary Tests. A group of six letters were written which spelled some nonsense word. These letters were rewritten three times omitting a different single letter each time. The subject is to write the missing letter in the blank space opposite it. For example, KESTYA: Kesta \_\_, Ksty<sup>a</sup> \_\_, Estya \_\_.

"The Digit-Symbol Test or substitution test," Wechsler states, "is one of the oldest and best established of all psychological tests. It is to be found in a large variety of intelligence scales, and its wide popularity is fully merited. The subject is required to associate certain symbols with certain other symbols, and the speed and accuracy with which he does it serves as a measure of his intellectual ability." (9)

The Vowel-Consonant Test applies a principle embodied in the Kuhlman-Anderson Intelligence Tests and the principle of following directions often used in intelligence tests. The subject is to make a mark through or near all vowels and double consonants in written material—underline single a's, encircle single i's, draw a line above single e's, dot the single o's, draw a line through single u's, put an "x" over joining double vowels, put a zero (0) under unlike joining vowels, and enclose all double, joining consonants in a circle.

The Parentheses Marking Test employs a combination of the principles used in the Digit-Symbol Test and in following-directions tests. However, in this case the exercise tends to be more concrete for the subject is to do something to the symbol. Ten sets of parentheses were given, one being without an enclosure and the other nine having some mark or symbol enclosed—( ), ( ' ), ( . ), ( 1 ), ( : ), ( / ), ( - ), ( " ), ( \* ). The subject is only required to draw a straight line, vertical or horizontal, through or near a part of the parentheses or symbol. One immediately sees that this exercise follows some concrete logic and tends to be learned easily, *e. g.*, drawing a line under the period, drawing a line over the apostrophe, dividing the colon, crossing the diagonal, drawing a line under the quotation mark, and drawing a vertical line above the dash.

The Reversed Type Test represents the acquisition of new habits with familiar material, *i. e.*, reading ordinary type which has been reversed and which reads from right to left on the page. This exercise is comparable to the mirror reading test used by Johnson. Simple statements asked a question or told the subject what to write in the space opposite it. This was a logical method of ascertaining whether or not the subject had read the statement in group testing.

**TABLE I**

**THE MEANS AND STANDARD DEVIATIONS OF INTELLIGENCE QUOTIENTS, SCHOOL MARKS AND THE LEARNING TEST SCORES OF 41 ELEVENTH GRADE PUPILS**

Item	Mean	Standard Deviation
Intelligence Quotient	88.46	8.62
School Marks	20.89	5.38
Total Score	664.62	144.55
Digit-Symbol	170.61	42.08
Vowel-Consonant	239.18	77.77
Parentheses Marking	126.26	38.76
Reversed Type	48.23	29.61
Letter Observation	81.79	14.47

Procedure

The Intelligence Quotients and School marks were secured from the pupil records. The former were determined by the Otis Quick-Scoring Mental Ability Tests, Gamma Test, Form AM, which had been administered to the group about a month prior to this testing. School marks for the first semester were converted into numerical values by computing averages for each pupil based on the following scale: A is 40; A-, 38; B+, 32; B, 30; B-, 28; C+, 22; C, 20; C-, 18; D+, 12; D, 10; D-, 8; F, 2.

The mean Intelligence Quotient was 88.46 with a standard deviation of 8.62. The mean school mark was 20.89, the equivalent of a "C" letter mark, with a standard deviation of 5.38. These means are shown in Table I.

The test was administered to a group of boys and girls before administering it to the subjects whose performance provided data for this study. The results of the former are not included in this report.

The subjects used in the study were 41 eleventh grade pupils, 23 girls and 18 boys, whose ages ranged from 15 years and 10 months to 18 years and 10 months. They live in and near a city in Delaware.

The test was administered to the pupils in two groups in a single morning by the writer. Each group received the same directions and there was no lapse of time between the testings to permit the passing of information. Each group was told that the test was simply a group of exercises which required no information learned in school and that care had been taken so that everyone would have an equal opportunity to make high scores and that it was, therefore, necessary for everyone to begin and to stop working at the same time. Time was kept by the use of a watch and an electric clock; both of which had second hands on them. Five minutes were allowed for the Letter Observation Test, 5 for the Digit-Symbol, 10 for the Vowel-Consonant, 5 for the Parentheses Marking and 10 for the Reversed Type Tests. Directions were read aloud to the pupils while they read silently from the booklet for each subtest. Instructions for each test except the Reversed Type Test were written on the blackboard, but were erased before the pupils began working. Practice was given before each subtest.

In giving instructions for the Reversed Type Test each letter of the alphabet was pronounced aloud by the examiner while the pupils followed on the booklets upon which were written both the usual and reversed letters of the alphabet. Pupils were told that saying the statements softly to themselves might help them to grasp the meaning more quickly.

A minute break was given after the first three exercises in which the pupils were asked to stand and "take a stretch." In order to measure the speed of all pupils, no one was able to finish any test within the allotted time.

TABLE II

CORRELATIONS BETWEEN SUBTESTS AND TOTAL TEST SCORES  
OF 41 ELEVENTH GRADE PUPILS

	Digit-Symbol	Vowel-Consonant	Parentheses Marking	Reversed Type	Whole Test
Letter Observation	.37 $\pm$ .13	.25 $\pm$ .14	.25 $\pm$ .14	.33 $\pm$ .14	.48 $\pm$ .12
Digit-Symbol		.27 $\pm$ .14	.14 $\pm$ .15	-.13 $\pm$ .15	.49 $\pm$ .12
Vowel-Consonant			.58 $\pm$ .10	.33 $\pm$ .13	.88 $\pm$ .03
Parentheses Marking				.53 $\pm$ .11	.79 $\pm$ .06
Reversed Type					.51 $\pm$ .12

Scoring the Tests

The tests were scored twice by the writer and throughout the first four exercises, pupils received one point for each correct response. On the Reversed Type Test, 3 points were given for each correct response. Sometimes the pupil's response indicated that he had read the statement, but only poorly. For example, opposite the statement "Write seven." a pupil wrote "aeven" for which he received 2 points. One pupil received one point for writing "brown" opposite "What color is snow?" Another pupil received 2 points for writing "Sunday" opposite "What day comes before Saturday?" The writer recognizes the weakness of this subtest in that the judgment of the scorer was used. However, the number of responses requiring judgment was very small.

It is noteworthy that pupils tended to score consistently high or consistently low throughout the entire test.

The means and standard deviations were computed for the total scores and for each subtest score. The mean total score was 664.62 with a standard deviation of 144.55. The mean of the subtests scores ranged from 81.79 with a standard deviation of 14.47 for the Letter Observation Test to 170.61 with a standard deviation of 42.08 for the Digit-Symbol Test. These means and standard deviations are shown in Table I.

Analysis of the Data

The standard deviations indicate that there is wide variation among the scores. The scores on the Reversed Type Test were most variable and tend to be bimodal when shown graphically. Correlations were computed for the subtests, the total, and subtests scores. Table II shows these correlations and standard errors. The correlations with their standard errors between the subtests ranged from  $-.12 \pm .15$  (Digit-Symbol and Reversed Type) to  $.58 \pm .10$  (Vowel-Consonant and Parentheses Marking). For the most part, these correlations tend to be low and therefore, show that the tests tended to measure different learning skills. The correlations between the subtests and total test scores are rather high for they ranged from  $.48 \pm .12$  with the Letter Observation Test to  $.88 \pm .03$  with the Vowel-Consonant Test.

The correlations between Intelligence Quotients, school marks, and test scores are rather low and are shown in Table III. These correlations and standard errors are  $.19 \pm .15$  between the learning tests and Intelligence Quotients and  $.16 \pm .15$  between school marks and learning test scores. The correlations between the subtests, Intelligence Quotients and school marks are uniformly low, ranging from  $.00 \pm .15$  for Intelligence Quotients and Digit-Symbol to  $.33 \pm .14$  for Intelligence Quotients and Letter Observation. These correlations and standard

TABLE III

CORRELATIONS BETWEEN INTELLIGENCE QUOTIENTS,  
 TEST SCORES, AND SCHOOL MARKS OF 41 ELEVENTH  
 GRADE PUPILS

Test	Intelligence Quotients	School Marks
Total Test	.19+.15	.16+.15
Parentheses Marking	.06+.16	.15+.15
Vowel-Consonant	-.14+.15	.13+.15
Digit-Symbol	.00+.15	-.02+.16
Letter Observation	.33+.14	.30+.14
Reversed Type	.28+.14	.14+.15

errors are also shown in Table III.

The correlation with its standard error between school marks and Intelligence Quotients of the group was found to be .54+.11.

The standard errors indicate in some instances, especially where the correlations are low, that the correlations would tend to be more reliable were they computed from scores made by a much larger group of pupils. This factor is regarded in the recommendation.

### Conclusions

If the learning tests employed in the study measured the ability to learn with a fair degree of reliability, it is reasonable to conclude that

1. The low correlations between subtests indicate that the tests were measuring different learning skills.
2. The low correlations indicate that there is little relationship between the ability to learn and intelligence as measured by intelligence tests.
3. The low correlations between school marks and the ability to learn indicate that factors other than the ability to learn operate to influence school marks.

### Recommendation

It is recommended that an instrument be constructed which will measure the innate ability to learn, eliminating as much as possible the influences of schooling and cultural backgrounds, and administered to a large and representative group of pupils and correlate the scores with those made on a group of the now widely used group intelligence tests.

### BIBLIOGRAPHY

1. Boynton, Paul L., Intelligence, Its Manifestations and Measurement. New York: D. Appleton and Company, 1913. 466 pp.
2. Freeman, Frank S., Theory and Practice of Psychological Testing. New York: Henry Holt and Company, 1950, 518 pp.
3. Greene, Harry A., Albert N. Jorgensen, and J. Raymond Gerberich, Measurement and Evaluation in the Secondary School. New York: Longmans, Green and Company, 1946, 670 pp.
4. Johnson, O. J., "A Study of the Relation Between Ability to Learn and Intelligence as Measured by Tests," Journal of Educational Psychology, 14: 540-44, December, 1923.
5. Jordan, A. M., "The Validation of Intelligence Tests," Journal of Educational Psychology, 14: 348-66, September, 1923.

6. Mursell, James L., Psychological Testing. New York: Longmans, Green and Company, 1947, 449 pp.
7. Peterson, Joseph, "Learning When Frequency and Recency Factors Are Negative," Journal of Experimental Psychology, 5: 270-300, August, 1922.
8. Pintner, Rudolph, Intelligence Testing Methods and Results. New York: Henry Holt and Company, 1935, 555 pp.
9. Wechsler, David, The Measurement of Adult Intelligence. Baltimore: The Williams and Wilkins Company, 1944, 258 pp.

# AN EXPERIMENT IN THE CONTINUITY-TYPE QUESTION

R. R. ASHBURN  
JANICE HILL BRADSHAW  
West Virginia University

THE CONTINUITY, rearrangement, or sequence question is one in which the student is asked to rearrange in the proper order a number of similar or related items. For instance, in history the student may be asked to arrange certain events in chronological order, or in geography, to arrange certain countries in order of size. This type of question is considered by many teachers a valuable testing device in those phases of learning in which it is desirable to test a sense of relationships rather than exact factual knowledge.

The scoring of a question of this type is rather complex since a fair score allows for different degrees of accuracy. Many articles have been written concerning scoring devices, but no agreement as to the best method has been reached. One writer, Worcester (1), has come to the conclusion that the only fair method procedure is not to use this type of question at all, since injustice is done by all of the scoring methods. His evidence, however, is based on imaginary cases and may not represent real situations.

The present writers have conducted two experiments to check and compare the various scoring methods proposed for the continuity question by controlling the knowledge of those taking the test.

A list of ten fictitious countries was made and an imaginary area assigned to each:

Bavoc	2,800,000 square miles
Vandalia	1,500,000 square miles
Generia	980,000 square miles
Wonderland	800,000 square miles
Oz	700,000 square miles
Electricia	680,000 square miles
Westerland	660,000 square miles
Mu	70,000 square miles
Lavania	8,500 square miles
Utopia	600 square miles

The list of countries was locked up and the following test was given in four sections of Humanities 2, to fifty-seven students:

## EXPERIMENT IN SCORING

You are invited to assist in a series of experiments to determine the accuracy of scoring examinations. This is the first of the series.

### EXPERIMENT I

Imagine yourself in the following situation.

You have not had time to do the assigned readings for one of your classes. The teacher surprises you by giving a test on this material. Since you know none of the correct answers to the questions, you decide to guess at them, thinking that perhaps you will get a few right. The test the teacher gives is as follows:

Arrange the following countries in the order of their size by placing the order number in the blank at the right. That is, put the figure (1) after the largest country, the figure (2) after the next largest, etc. Do not omit any countries, for the question cannot be graded unless every blank is filled. If you get as many as two countries in the proper order, you will receive some credit.

Wonderland	( )
Vandalia	( )
Oz	( )
Utopia	( )
Bavoc	( )
Electricia	( )
Westerland	( )
Generia	( )
Mu	( )
Lavania	( )

The answers were scored by ten different methods, all suggested by earlier writers.

Wilson (2) first proposed basing the student's score on the sum of the differences between the order given by the student and the true order of the several items. This principle, with variations of numerical value and scoring devices, provided the methods of Sangren and Woody (3), Odell (4), Richardson, Russell, Stalnaker and Thurstone (5), and Rinsland (6). Odell (7) suggested three theories for assigning maximum and minimum scores:

1. Score of zero should be assigned to the complete reversal of the correct order, which is the worst possible answer. The score for chance order should be one-half of the maximum score.

2. Chance arrangement should be scored zero, since this order is indicative of no knowledge. Score given a complete reversal should be negative.

3. Score of zero should be assigned to all answers in which there is a negative correlation with the correct answer as well as to those answers which exhibit chance arrangement.

Wilson's method, varied by applying to it each of the three theories of maximum scores, provided our methods numbered I, II, and III.

Methods IV, V and VI are based on the rank correlation of answers to true order as proposed by Cureton and Dunlap (8) and differ only in maximum and minimum score assignments.

Methods VII and VIII are based on the system by which a student is given a certain number of points for each item arranged anywhere below the item which should immediately precede it, as first proposed by Nesmith (9) and modified by Lang (10), Wilson (11) and Rosander (12). These two methods differ only in assignment of minimum and maximum score.

Method IX is based on the number of exact sequences as described by Montgomery (13).

Method X used Montgomery's formula (13) for finding the sequence-rank grade, which is merely an average of the percentage of correct sequences with the percentage grade determined by the sum of the differences.

All scores were converted to a scale of 100. The ten methods, then, are:

- I. Based on sum of possible differences between correct order and the order given by students. Score of zero assigned to complete reversal of correct order.
- II. Based on sum of possible differences between correct order and the order given by students. Answer of one-half of maximum sum of differences receives score of zero. Answers of more than one-half of maximum sum of differences receive negative scores.
- III. Based on sum of possible differences between correct order and the order given by students. Answers of one-half or more of maximum sum of differences receive score of zero.
- IV. Based on rank correlation or sum of the square of possible differences. Chance arrangement ( $1/2$  possible  $\Sigma d^2$ ) receives score of 50. Perfect negative correlation receives mark of zero.
- V. Based on rank correlation or sum of the square of possible differences. Chance arrangement ( $1/2$  possible  $\Sigma d^2$ ) receives score of zero. Negative correlations receive negative scores.

- VI. Based on rank correlation or sum of the squares of possible differences. Chance arrangement ( $1/2$  possible  $\Sigma d^2$ ) and all negative correlations receive score of zero.
- VII. Based on number of order-relationships. Score of zero assigned to complete reversal of correct order.
- VIII. Based on number of order-relationships. Score of zero assigned to answers of one-half or less of all possible order relationships.
- IX. Based on number of exact order sequences.
- X. Average of score based on sum of differences, considering complete reversal as worst possible answer, and score based on number of exact sequences (i.e. average of Methods I and IX).

Since the students had no information about the assumed sizes of the ten countries, it is obvious that all should receive a score of zero. All scores other than zero must be the result of chance or of the inadequacies of the scoring system. The actual scores received are tabulated in Table I.

The absurdity of negative scores is seen immediately from the number of such scores resulting from Methods II and V. Although the students had no information and hence no possible misinformation, Method II gives fifty-three negative scores and Method V gives thirty-two. Method II would assign a score of -92 to three papers.

Equally absurd are the results of all methods which assume that a complete reversal of the correct order is the worst possible answer and should receive a grade of zero, namely Methods I, IV, VII, and X. The average grades assigned to the fifty-seven papers by these methods are I, 38; IV, 47; VII, 47; X, 21. Method IX based on number of correct sequences, gives the proper score of zero to twenty-two students, but a score of more than 10 to thirty-five students. The remaining three methods (III, VI, VIII), all of which assign a score of zero to one-half, or more, of possible errors, represent more nearly the actual situations. Method VIII gives a score of zero to thirty papers, with highest score, 42, on two papers. Method VI gives score of zero to thirty-three papers, but with one score of 56. Method III gives a score of zero to fifty-three of the papers, with scores of 4, 12, 12, and 12 to the other four.

The distribution of the fifty-seven scores according to the ten methods does not tell the whole story. One might judge from the table that one method merely assigns a lower or higher grade in some fixed proportion. Such is not the case. On one paper the difference between scores assigned by two methods is 97. The scores for this paper by the different methods were: I, 24; II, -52; III, 0; IV, 41; V, -19; VI, 0; VII, 47; VIII, 0; IX, 22; X, 23.

TABLE I

## DISTRIBUTION OF SCORES IN EXPERIMENT I BY SCORING METHODS

Scores	Number of Scores by Method									
	I	II	III	IV	V	VI	VII	VIII	IX	X
-100 to -91	3									6
- 90 to -81	1									19
- 80 to -71	1									25
- 70 to -61										7
- 60 to -51	10									
- 50 to -41	5									
- 40 to -31	9									
- 30 to -21	5									
- 20 to -11	15									
- 10 to - 1	4									
0	53	1	4	3						
1 to 10		10	6	3						
11 to 20		20	6	9						
21 to 30		30	10	6						
31 to 40			23	3						
41 to 50			10	14						
51 to 60			10	2						
61 to 70			11	2						
71 to 80			10	14						
81 to 90			2	2						
91 to 100			7	7						

On the basis of our first experiment, we conclude that Method III is the only one of the ten that approaches a fair grade in the case of a student who is completely ignorant of the facts.

To further test the accuracy of these ten methods, we asked fifty-three students in the same four sections of Humanities 2 to take part in the following experiment:

### EXPERIMENT IN SCORING

This is the second in the series of experiments to determine the accuracy of scoring examinations. You are again invited to participate.

### EXPERIMENT II

Imagine yourself in the following situation.

On mid-semester examination the teacher gives the same question you had on the surprise test. However, you have now done your assigned readings and remember the following facts about the countries concerned:

1. Bavoc is the largest of all and Utopia is the smallest.
2. Lavana boasts that its culture is superior to that of Mu, even if the latter is greater in size.
3. Wonderland, which is a subdivision of Generia, once attacked its smaller neighbor, Oz.
4. Electricia became slightly smaller than Oz when it lost a tenth of its area in the separation of its colonies of Lavana and Mu.

Now, do your best on the question!

(The question used in Experiment I is repeated here)

The answers were scored by the same ten methods as in the first experiment.

The information given is roughly equivalent to that sometimes possessed by students under similar circumstances. There may be some difference of opinion as to the exact score that should be allowed for the given amount of knowledge. However, since the students have no knowledge of two of the countries, the score should certainly not exceed 80. Since the information given, if correctly applied, enables them to establish thirty-two or 71% of the forty-five order-relationships, we may say that the score should not be less than 71. All scores above 80 or below 71 must be the result of (1) chance, (2) inadequacies of the scoring system, or (3) failure to apply the information supplied. The

TABLE II

DISTRIBUTION OF SCORES IN EXPERIMENT II BY SCORING METHODS

Scores	Number of Scores by Methods:									
	I	II	III	IV	V	VI	VII	VIII	IX	X
-100 to -91	0	1 to 10	(6)	(1)	(1)	(1)	(1)	(1)	(1)	(1)
- 90 to -81	11 to 20	(8)	(8)	(1)	(1)	(1)	(1)	(4)	(4)	(1)
- 80 to -71	21 to 30	(3)	(3)	(2)	(1)	(12)	(12)	(10)	(5)	(2)
- 70 to -61	31 to 40	(2)	(2)	16	16	(1)	(1)	(1)	(1)	(7)
- 60 to -51	41 to 50	(5)	2	(1)	2	(1)	(2)	(2)	(2)	(6)
- 50 to -41	51 to 60	(11)	4	(1)	4	(1)	2	(4)	6	(4)
- 40 to -31	61 to 70	(2)	4	4	(13)	16	(2)	(11)	12	(3)
- 30 to -21	71 to 80	(2)	2	2	(2)	1	(1)	6	(4)	19
- 20 to -11	81 to 90	(3)	2	2	(2)	6	(1)	6	(1)	3
- 10 to -1	91 to 100	(2)	2	2	(2)	7	(1)	16	(1)	8

extent to which the students fail to apply the information is a measure of their ability to fit their knowledge to the framework of the test and not a measure of their knowledge. The actual scores received are tabulated in Table II. Numbers in parentheses represent scores of students who failed to apply correctly the information supplied.

Although all the students had the same knowledge of the facts, the distribution of scores is as absurd as in the first experiment. One student arranged all the items in correct order and hence received a score of 100 by all methods. Only in the case of a perfect paper do the ten methods agree on the score! The distribution of scores for the twenty-one students (38% of the group) who failed to use correctly the information given is obviously due to chance. For these students, this question has not been a test of knowledge. Only Method VII gives a score of 71 or more to all students making use of the information supplied, but this method assigns scores above 80 to twenty-six (or 81%) of the thirty-two papers. Method IX gives scores of less than 71 to all papers except the one perfect paper. None assigns a majority of scores between 71 and 80. The variation of score by different methods on the same paper is as much as 97 in the case of a student who did not apply correctly the information. This paper received a score of -68 by Method II and 29 by Method VII. Of the papers using the information correctly, one received a score of 60 by Method III and 93 by Method IV, another 73 by Method VI and 98 by Method VII.

From the results of these two experiments we conclude:

1. None of the existing methods of scoring the continuity-type question are sufficiently accurate to justify basing any part of a student's grade on such a question.
2. The continuity-type question should be discarded until or unless there be devised a system that will compensate for the chance variations of scores.

## BIBLIOGRAPHY

1. D. A. Worcester, "Still Further Comments on the Scoring of the Continuity Test", School Review, XXXVIII, June, 1930, pp. 462-466.
2. H. E. Wilson, "The Continuity Test in History-Teaching", School Review, XXXIV, Nov., 1926, pp. 679-684.
3. P. V. Sangren and C. Woody, Sangren-Woody Reading Test, Manual of Directions, World Book Company, Yonkers-on-Hudson, New York 1927, p. 6.
4. C. W. Odell, Traditional Examinations and New-Type Test, Century Co., New York, 1928, pp. 406-408.
5. M. W. Richardson, J. T. Russell, J. M. Stalnaker, L. L. Thurstone, Manual of Examination Methods, University of Chicago, Chicago, Ill., 1933, p. 104.
6. H. D. Rinsland, Constructing Tests and Grading, Prentice Hall, New York, 1938, pp. 144-145.
7. C. W. Odell, "The Scoring of Continuity or Rearrangement Tests", Journal of Educational Psychology, XXXV, Sept., 1944, pp. 352-356.
8. E. E. Cureton and J. W. Dunlap, "Scoring the Rearrangement or Continuity Test", School Review, XXXVIII, Oct., 1930, pp. 613-616.
9. R. W. Nesmith, "Scoring the Continuity Test", School Review, XXXVII, Dec., 1929, pp. 764-766.
10. A. R. Lang, Modern Methods in Written Examinations, Houghton Mifflin Co., Cambridge, Mass., 1930, pp. 194-199.
11. H. E. Wilson, "Further Comments on the Scoring of Continuity Tests", School Review, XXXVIII, Feb., 1930, pp. 115-123.
12. A. C. Rosander, "A Simple Method of Scoring and Interpreting Sequential Responses", Journal of Educational Research, XXXVI, Nov., 1942, pp. 168-177.
13. E. A. Montgomery, "Construction of Sequence Rank Tests", Journal of Educational Research, XXXIX, March, 1946, pp. 523-527.

# AN INVESTIGATION OF CERTAIN FACTORS RELATING TO THE NATURE OF CHILDREN WITH FUNCTIONAL DEFECTS OF ARTICULATION

OLIVER W. NELSON  
University of Washington  
Seattle 5, Washington

THIS STUDY was concerned with an effort to gain a better understanding of children with functional articulatory defects of speech by an investigation of (1) personal adjustment and social behavior, (2) school progress, (3) school attendance, (4) health, (5) intelligence and (6) leisure-time interests and activities.

Inclusion of the foregoing factors in the study was motivated by the following tentative observations: First, there is conflicting evidence in the current literature regarding social adjustment of such children; second, there is a possibility of invalidity of previous findings regarding mental abilities of these children in cases of scores derived from instruments possessing a prominent verbal factor; third, there is inconclusive evidence from previous investigations regarding the health of speech defectives; fourth, there is apparently no objective evidence regarding the school attendance and leisure-time interests and activities of such children, and fifth, all six factors are pertinent to a study of the nature of the speech defective child.

Representative of inquiries into the nature of the social adjustment behavior of speech defectives are the investigations reported by Van Riper (9), Berry and Eisenson (2), Reid (6), McAllister (5) and Beckey (1). That the speech defective may present a distinct personality portrait is inferred from descriptions of such individuals offered by most authors of textbooks in speech correction. However, certain controlled studies of elementary school children tend to cast some doubt on the assumption that the speech defective differs significantly from the normal speaker in emotional and social adjustment. Reid and McAllister, using personality inventories with elementary school children, found no significant difference between personal and social adjustment scores of speech defectives and those of normal speakers with whom they were compared. Likewise, Beckey, employing a case history approach, compared fifty speech defectives with controls and failed to find any significant difference between the groups in frequency of behavior problems.

However, the validity of the findings of most of these studies was questioned, since it was not known in many cases whether prior to the investigation the speech defectives had received any speech training or general guidance. Such training would most certainly tend to invalidate the findings of comparative studies aimed at personality analysis.

Excluding those with defective intelligence, children with articulatory problems, as a group, are generally reported to be slightly inferior in mental capacity to the general population. Typical of the studies reporting such findings is that of Carrell's investigation (3) of 1174 school children in Mooseheart, Illinois. However, in arriving at this finding the majority of the investigators have employed testing instruments that possess a prominent verbal factor. This procedure may be challenged on the ground that such tests do not provide an equitable basis for pairing subjects who possess linguistic disabilities with those who do not have such limitations. The present study sought to meet this objection by using a testing instrument limited to non-language factors.

Insufficiency of confirming evidence causes the writer to accept as tentative Beckey's findings (1) based on medical reports that poor health was significantly more common among speech defectives than among the controls examined.

No objective evidence could be found in the literature either to support or refute the assumption that children retarded in speech are more frequently absent from school or have fewer leisure-time interests and activities than normal speakers of comparable age, grade and intelligence.

On the basis of these observations there appeared ample justification for further study of the nature of children with functional defects of articulation in terms of factors and methods of approach selected for the present investigation.

### Procedures

Subjects for this experiment were drawn from a population of 1,659 pupils in the third, fifth and seventh grades of seven elementary schools in suburban Seattle, Washington. No previous organized program of remedial work had been offered the children in this school system. There were, therefore, no "effects" of special instruction to be taken into consideration in evaluating the findings. The school district population represented a good cross-section of various socio-economic levels. Third, fifth and seventh grades were selected as three important levels of pupil maturity and were believed to be representative of the entire elementary school population.

Subjects were selected through a combination of survey and personal interview methods. Standard procedure for testing children with

articulatory problems was followed. In the course of the examination all children found to have hearing losses and organic deviations which might be responsible for the speech disorder were eliminated from the final experimental groups. Data regarding mental capacity were based on scores obtained from non-language factors of the California Test of Mental Maturity.

In this study the matched-pair technique was employed. Children for control groups were selected from the same grades and schools which supplied the experimental subjects. Equating was accomplished on the bases of grade, sex and intelligence quotients derived from non-verbal factors. In the equating process no greater than a five-point differential between paired intelligence scores was permitted.

As previously indicated, six different factors were examined in the study. These were: (1) intelligence, (2) personal and social adjustment, (3) scholastic success as indicated by school marks and age-grade level, (4) school attendance, (5) health and (6) leisure-time interests and activities.

Data for the first and sixth factors were obtained from the children's responses to questions in the California Test of Personality and teachers' ratings of these children on the Haggerty-Olson-Wickman Behavior Rating Schedule. Data for the factors of educational progress and school attendance were secured from pupils' permanent records and from personal interviews with teachers and school principals. Information regarding health was secured from reports of surgeons and physicians as indicated on pupil health and medical cards and from interviews with school nurses. For statistical purposes these data were given a numerical value by rating the pupils on a three-point scale.

In employing over-ageness as a measure of school progress two factors were considered: (1) number of children beyond the normal age range for their grade and (2) number of months of over-ageness. To allow for normal variations in ages found among children entering school at the first grade a differential of six months was permitted above and below the standard age.

Tabulated scores earned by the defective speech groups and their controls in each of the six factors introduced in the study were subjected to standard statistical treatment. The significance of the mean differences between experimental and control groups was determined by the T test of significance carried to the 5 per cent level of confidence. The findings, which will now be reviewed, may, therefore, be considered to possess an acceptable degree of reliability.

#### Summary of Findings

Thirty-five children from a population of 1,659 third, fifth and

seventh grade children were found to have functional disorders of articulation, a number which represents a percentage of 2.1 for the three age levels in question. This figure is slightly lower than that reported by other investigators such as Loutitt (4), Tanberg (7), Carrell (3) and Yedinack. (8) The difference may be attributed to the facts that first and second grade pupils were not included here and that only functional disorders were investigated.

Distribution by sex showed boys outnumbered girls two to one in the third grade but only four to three at the seventh grade level. It would seem that on the basis of this survey that the influences of maturation and school socialization, which normally serve to reduce certain speech defects, may not have operated as favorably for girls as for the boys studied here. Why this may be true was beyond the limits of this investigation to predict.

Findings of this study indicate that the children with functional defects of articulation did not differ fundamentally from the general population in intelligence when tested by non-language factors of a group intelligence test. The mean intelligence quotients of the experimental groups and the school population were 100.8 and 105.6 respectively. These scores did not yield a statistically significant difference. This finding is not entirely in accord with the conclusions of Carrell (3) and Berry and Eisenson (2) who reported intelligence of their subjects to be significantly lower than the general population from which they were selected. Further investigation might show that the verbal factor in intelligence tests produced the element of difference between intelligence scores found by this study and those reported by other investigators.

When only the findings of the California Test of Personality were considered, no support could be found for the contention that children with functional disorders of articulation are significantly more poorly adjusted than normal speakers of comparable age. It was interesting to note, however, that the teacher ratings in the Haggerty-Olson-Wickman scales showed that, taken as a whole, the experimental cases were considered to have less desirable behavior than the control group. The significance of this discrepancy is not clear, although it might be interpreted as raising some doubt regarding the validity of one of the two scales.

The examination of scholastic success as indicated by school marks confirms the results of Carrell's study (3) on this point in finding that the experimental subjects were not succeeding as well in school as were the controls with normal speech. The speech defectives averaged the equivalent of one letter grade lower than normal speakers. This investigation also found that while experimental subjects in the third and fifth grades were not significantly more over age for their grade than the controls, subjects in the seventh grade averaged 12 months

more over age than their controls, which fact suggests that special problems of learning may be associated with defective articulation.

The foregoing conclusion is educationally significant in view of the finding that children with functional disorders of articulation attended school as regularly, had generally as good health and possessed as many and as varied leisure-time interests and activities as normal speakers of their particular grade, sex and mental ability.

### Inferences

Findings of this investigation which reveal both differences and similarities between the speech defective and control groups may be helpful in furnishing a clearer understanding of the nature of children with defective articulation and as such may be instrumental in providing this type of child with more appropriate guidance and instruction.

Results of the study would seem to justify a number of specific inferences: First, remedial attention to articulatory speech disorders at the earliest feasible opportunity is desirable. From a practical point of view this would mean, in most instances, providing correction for children in the second grade and above. Second, in view of the questions regarding the validity of the personality inventory and behavior rating scales raised by this study, it is the writer's opinion that a continuing study of the personality structure of the speech defective child should be made by use of other instruments such as the various projective techniques.

Third, since nothing in this investigation indicated the speech deviate to be significantly different from the general school population with regard to intelligence, and since it appears that as a group they do, nevertheless, tend to have lower scholastic achievement than normal speakers in the same grade level and intelligence, it may be necessary and advisable to provide such children with special individual assistance in school subjects if they are to realize their actual potential. The fact that many of them appear to be over age for their grade suggests the need of adapting materials of instruction to their level of physical and social maturity.

Fourth, the fact that children with articulatory speech defects appear to have as many and as varied special interests and activities as do normal speaking children suggests a fortunate circumstance for planning varied and stimulating educational experiences for such children.

## BIBLIOGRAPHY

1. Beckey, Ruth E., "A Study of Certain Factors Related to Retardation in Speech," Journal of Speech Disorders, 7: 223-249, September, 1942.
2. Berry, Mildred and Jon Eisenson, The Defective in Speech. (New York: F. S. Crofts and Co.), pp. 65-69.
3. Carrell, James, "A Comparative Study of Speech Defective Children," Archives of Speech, 1: 179-203, June, 1936.
4. Loutitt, G. M., and E. C. Ralls, "A Survey of Speech Defects Among the Public Schools of Indiana," Journal of Speech Disorders, 1: 73-80, September, 1936.
5. McAllister, Margaret, "A Study of the Relation Between Defects of Articulation in Speech and Emotional Instability in Elementary School Children," Unpublished Master's Thesis, University of Washington, Seattle, 1948.
6. Reid, Gladys, "The Etiology and Nature of Functional Articulatory Defects in Elementary School Children," Journal of Speech Disorders, 12: 143-150, June, 1947.
7. Tanberg, M. C., "A Study of the Incidence of Speech Defects Among School Children in the State of Washington," Unpublished Master's Thesis, University of Washington, 1937.
8. Yedinack, Jeannette G., "A Study of the Linguistic Functioning of Children with Articulatory and Reading Disabilities," Unpublished Doctoral Dissertation, Fordham University, 1947.
9. Van Riper, C., Speech Correction, Principles and Methods. New York: Prentice Hall.

# IMPLICATIONS OF PRIVATE AND PAROCHIAL SCHOOL POPULATIONS\*

RALPH C. GEIGLE

Oakmont (Allegheny County), Pennsylvania

OCCASIONALLY WE see or hear comment to the effect that private and parochial school enrollments are increasing in this country. About twenty-five years ago when first research was directed to educational finance on the inter-state level, it was conjectured that existing variabilities among the states might be partially explained by the presence of varying percentages of children of school age attending private and parochial schools. As early as 1920, Ayres commented that "The state having a relatively large proportion of private and parochial school children is rewarded in that such a state ought to make a better showing in the matter of expenditure for the pupils who are enrolled in public schools than it could if it had to bear the whole burden of expense." Norton in 1926 submitted that the "factor of parochial and private school attendance is of sufficient significance in some states to justify special consideration." The conditions of the twenties and their significance suggest an appraisal of private and parochial status in this our more contemporary period.

Have private and parochial school enrollments increased since the twenties? Are there variations among the several states in the percentages of the school-age population attending the private and parochial schools? If so, are such enrollments of such proportions as to seriously affect the state fiscal indexes?

Investigation has revealed that after 1935 (9.05 percent) there has been a consistent trend in the direction of increasing percentages of children attending private and parochial schools with the 1947-48 school term finding well over 11 percent so enrolled. From the viewpoint of diversity in the percentages attending private and parochial schools in the several states, a cursory examination of the data collected convincingly showed that considerable variation exists. In 1929, New Hampshire had more than twenty-six percent attending private and parochial schools while South Carolina, Oklahoma, and Nevada each had less than one percent. In 1935, New Hampshire's percentage had increased well over twenty-eight percent while twelve states each still had less than two percent so enrolled. In general, this diversity has continued to prevail down to our most recent years. In 1947, Rhode

\*An aspect of a dissertation submitted to the Faculty of the School of Education of the George Washington University in partial satisfaction of the requirements for the Degree of Doctor of Education.

Island had over twenty-eight percent enrolled in private and parochial schools while thirteen states each still had less than three percent so enrolled.

The continued diversity in the percentages enrolled in private and parochial schools coupled with still increasing enrollments justifies special consideration of the factor, certainly more than was warranted in the late twenties. If statistically interpretable, it remains to be seen how seriously this factor would affect fiscal indexes.

One approach to measuring the effect of private and parochial school enrollments would be to compare the efforts of the people as represented by its state to support public education with the efforts of the people to support public, private and parochial education. Such an adjustment was statistically possible only by assigning proper weight to the private and parochial school enrollments. This was accomplished by assuming that the economic cost per pupil enrolled in private and parochial schools is the same as that in public schools. In order to be as recent as possible, data for the 1947-48 school term were examined. Effort is defined as the percentage of total individual income payments expended for education. It is explanatory to note that the consideration of private and parochial expenditures would increase the index of effort in all states but obviously, in view of the diversities indicated above, not to the same degree.

For the nation as a whole the consideration of expenditures for private and parochial schools would increase the index of effort by 12.5 percent. Rhode Island and New Hampshire would have effort indexes increased by 38 and 34 percents respectively. The New England states as a whole would better their effort rating by 29 percent while the South Atlantic States would show but a 3 percent betterment in their index of effort.

It follows from the above that those states with high percentages of private and parochial school enrollments would gain places in the effort ranking while those states with low percentages would suffer correspondingly. When the states were ranked first on the basis of the percent of total individual income payments expended for public education certain acute shifts were noted. Illustrative of the contrary effects were seen in the state of Vermont, which stood to gain thirteen places, and Georgia, which stood to lose ten places.

It was sensed during the investigation that consideration of costs of private and parochial school enrollments as regards indexes of effort is especially comforting to the wealthy states. In fact, it is usually thought that private and parochial school enrollments might be responsible for the fact that wealthy states as New York and Rhode Island seem to devote smaller percentages of their fiscal abilities to public education than the average for the poorest states of the union.

Investigation revealed that the twelve wealthiest states (based on total income payments per child 5 to 17 years of age inclusive) in the 1947-48 school term had 16.29 percent of the school-age group enrolled in private and parochial schools while the twelve least wealthy states had but 3.61 percent so enrolled. This of course connotes that the wealthiest states stand to gain more face as regards efforts indexes by the consideration of the costs of private and parochial school enrollments. Statistically, the consideration of private and parochial school enrollments would increase the index of effort of the twelve wealthiest states by 19 percent while for the twelve least wealthy states but 4 percent.

There is yet another way in which one might interpret the effect of private and parochial school enrollments. This supplementary analysis is based on the three basic observations that, (a) State support in all states is restricted to public schools, (b) Federal aid as envisioned in our Federal aid bills would be distributed to the State authority, thus earmarking such funds for public schools, and (c) Federal aid bills in general have employed the number of children five to seventeen inclusive, in each state, as the "unit of need" in determining the apportionment to each state. With the above in mind, let us use the provisions of the Barden bill in determining the allotments to two states. To better illustrate the point to be made, we should apply these provisions to two states that are quite diverse in their percentages enrolled in private and parochial schools. To this end, Louisiana with more than fifteen percent and North Carolina with less than one percent enrolled in private and parochial schools in the 1947-48 school term are selected. The tabular data needed are submitted in Table I.

The Federal aid bill, as is true of all such Federal aid legislation, was designed along lines of the equalization principle, to help the less wealthy states more than the wealthy states. If the appropriation to these two states would have been as indicated in the table, the "equalization difference" would have been \$4.63 (\$23.54 - \$18.91). In other words, North Carolina would be entitled to an additional \$4.63 per school-age child because of her weaker financial status per school-age child. However, the number of school-age children, 5-17 inclusive, is only a theoretical burden as far as the state is concerned. The Federal allotment stands to be apportioned only over that portion of the school-age group that is enrolled in public schools, thus favoring, as far as the equalization is concerned, the public school age group in the more wealthy of the two states, Louisiana. Actually, as far as the school-age recipients of the aid are concerned, the "Equalization difference" between the two states is but \$2.11 (\$23.68 - \$21.57) per child rather than \$4.63. If one assumed that North Carolina must receive \$4.63 more per child than does Louisiana in order to properly

TABLE I

THE EFFECT OF PRIVATE AND PAROCHIAL SCHOOL ENROLLMENTS ON FEDERAL AID  
 IN ACCORDANCE WITH THE PROVISIONS OF THE BARDET BILL IN TWO STATES,  
 LOUISIANA AND NORTH CAROLINA, 1947-48

State	School age population	Private and parochial school enrollment	Amount payable to the state	Amount payable per child 5-17 inclusive	Amount payable per child not enrolled in private and parochial schools
	5-17 inclusive 1945	1945	c	d	e
1	2	3	4	5	6
North Carolina	973,401	5,520	\$22,920,000	\$23.54	\$23.68
Louisiana	607,705	75,184	11,490,000	18.91	21.57

<sup>a</sup>Source: U. S. Office of Education, Biennial Survey of Education, Statistics of State School Systems, 1945-46, Chapter II, (Washington, D. C.: Government Printing Office, 1949), p. 48.

<sup>b</sup>Ibid., p. 94.

<sup>c</sup>Source: Ralph C. Geigle, "The Relative Efforts of the States to Support Public Education" (Unpublished Ed D dissertation, School of Education, the George Washington University, 1950), p. 115.

<sup>d</sup>Figures in this column are computed by dividing figures in Column 4 by those in Column 2.

<sup>e</sup>Figures in Column 6 are computed by dividing figures in Column 4 by Column 2 less Column 3.

"equalize", the recognition of the private and parochial school enrollment would have the effect of nullifying by 53 percent the purpose of the Federal aid bill in these two states.<sup>1</sup>

It is seen that private and parochial school enrollments are increasing and that consideration of them does seriously affect fiscal indexes. That it does and would have a bearing on public school finance has been amply testified above. State school budgets reflect the presence of private and parochial school enrollments within the state and this in turn is reflected in the percentage of the State tax dollar required to support the public education program. Legally, and by the very nature of the case, hypothetically, every school-age child, 5-17 inclusive, is a ward of the State. Any child may return to the public schools from either a private or parochial school. Consideration of private and parochial school enrollments for the most part in determining indexes of effort under present circumstances must remain a theoretical consideration based on the hypothesis that if such children were in attendance in the public schools, the State would expend in their behalf a like amount of money per child as it now expends on the wards carried on the rolls of our public schools.

<sup>1</sup>Computed on the basis of the difference between the two "equalization differences" (\$4.63 - \$2.11).

# A TECHNIQUE FOR A STATISTICAL INTERPRETATION OF AN INTEREST SURVEY QUESTIONNAIRE

SAM S. BLANC  
East High School  
Denver, Colorado

AS PART OF a study carried on during the school year 1950-51,<sup>1</sup> the science interests of pupils in a sampling of junior high schools in Denver were surveyed. A questionnaire was constructed using the current objectives of science teaching as a basis. A large number of science learnings, understandings, and outcomes were reviewed in order to select items to be included in the questionnaire. These statements were submitted to a jury of science teachers, and to a group of pupils, for the selection of items most pertinent to the major objectives of science teaching. Using an average ranking for these two groups of choices, 205 items were chosen for inclusion on the final form of the questionnaire. This was administered to a heterogeneous sampling of seventh, eighth, and ninth grade pupils in three junior high schools. These schools were selected on the basis of the general economic levels from which the pupils were drawn to give a better picture of the average interests of pupils throughout the city.

Two classes on each of the grade levels in each school were selected for this survey. The pupils selected were in general science classes, since this study was concerned with interests in science. The total number of questionnaires returned from the three schools was 486. The questionnaire was administered to the pupils in the three schools within the same period of time. All the questionnaires were distributed, marked by the pupils, and returned for tabulation during January, 1951.

The matter of how to treat the results of this questionnaire was given considerable thought. Several factors complicated the interpretation of the actual markings in the Yes and No columns on the questionnaire. The desired results were to be in terms of three grade levels with a separate interpretation for the boys and girls. The number of cases at each grade level and the division between the boys and girls was different in each case. Therefore, it was felt

<sup>1</sup>Sam S. Blanc, "Audio-Visual Resources for the Teaching of Science in the Junior High School", unpublished Doctor's dissertation, University of Denver, Denver, Colorado, 1951.

that the raw scores could not be used directly. To obtain a more accurate interpretation, it was decided to employ the percentages of the total cases in each category of grouping.

Another factor to be considered was the fact that negative interests had to be considered as well as the positive interests. In other words, what was the meaning on a particular item if as many pupils checked the No column as the Yes column? Or, how significant were the results for a particular item if 50 cases checked the Yes column and 25 cases checked the No column? This type of tabulation certainly called for a different interpretation than one on which 50 cases checked the Yes column and only 15 checked the No column. How much more significant was the first case than the second? How the degree of interest shown for each item could be measured on a comparative scale for each group at each grade level was the problem that had to be solved in interpreting the meaning of the results of the questionnaire.

The first step necessary was the tabulation of the raw scores from the papers. The number of pupils at each grade level, boys and girls separately, placing a check in the Yes column and the number of pupils placing a check in the No column was tabulated for each item on the questionnaires. The checks placed in the Undecided column and the items omitted were not tabulated, since it was felt that if the pupil were uncertain about his reaction to a particular item, no conclusion for that answer could be drawn.

The second step was the determination of the percent of the total cases in each group which each score on the Yes and the No answers represented. To account for positive and negative interests in the interpretation of these results, the difference in percents for each item in each group was computed. An illustration of the percentage difference in answers on one item will make the steps in this computation clear:

Total number of returns for seventh grade boys	85
Number of checks in Yes column for item 1	23
Percent of seventh grade boys marking Yes	27.04
Number of checks in No column for item 1	39
Percent of seventh grade boys marking No	45.88
Actual difference in percents	-18.82

The difference in percent might be in favor of the Yes answers, thereby indicating a positive interest in that item, or the difference might be in favor of the No answers, thereby showing a lack of interest in that item. The greater the difference in percents, whether positive or negative, the more significant the conclusions drawn for that topic might be. If the difference in percents between the Yes and No scores

for an item were negligible, then no real conclusions as to interest or lack of interest in that item may be drawn. In a sense, this was equivalent to the statistical technique of assigning a +1 value to each Yes answer, a -1 value to each No answer, and a zero value to each Undecided or omitted answer.

To determine a point at which the difference in percent would have a statistical significance, the discussion on standard errors of proportions and percents in sampling was consulted in Lindquist.<sup>2</sup> It was found that the following formula was applicable in this situation:

$$\sigma_p = \sqrt{\frac{pq}{N}} \quad \text{op is the standard error in the percent}$$

$p$  is the percent in one category

$q$  is the percent in the other category

$N$  is the number of cases in the sample

To illustrate the use of this formula let us assume that on one particular item for the eighty-five seventh grade boys the number of Yes answers was 50 percent, and the number of No responses was 25 percent of the total cases in that group. Applying the formula we have,

$$\sigma_p = \sqrt{\frac{(0.50)(0.25)}{85}}$$

$$= \sqrt{\frac{0.1250}{85}}$$

$$= \sqrt{0.0015}$$

$$= .039$$

The critical ratio (C. R.) was determined by dividing the difference of the two percents by the standard error of the percent ( $\sigma_p$ ) computed above.

$$C. R. = \frac{p - q}{\sigma_p}$$

$$= \frac{.50 - .25}{.039}$$

$$= 6.41$$

<sup>2</sup>E. F. Lindquist, A First Course in Statistics (Boston: Houghton Mifflin Company, 1942), pp. 125-29.

In the table dealing with the minimum values of critical ratio required for significance at various levels of confidence in Lindquist,<sup>3</sup> it was found that the critical ratio of 6.41 far exceeded the 0.1 percent level of confidence. Hence, such a percent difference would be considered highly significant from a statistical point of view. However, for a study of this nature, it was felt that a critical ratio at the 0.1 percent level of confidence was too high for the type of conclusions which might be drawn. It was decided, therefore, to use the five percent level of confidence as a point above which the values derived in interpreting the responses to the items would be considered significant for this study.

In order to find this point at which the difference in percent would be considered of statistical significance for each of the groups of data, the standard error of the percent, using the above method, was determined for the boys and girls at each of the grade levels. These figures were found to be:

Seventh grade boys	.054
Seventh grade girls	.055
Eighth grade boys	.058
Eighth grade girls	.058
Ninth grade boys	.057
Ninth grade girls	.052

In the table on page 240 in Linquist<sup>4</sup> the critical ratio for the five percent level of confidence for a sample size of between 60 and 120 is given as 2.000. Any value of 2.000 or more as a critical ratio would, therefore, be considered statistically valid in this investigation. To find this point in the differences in percent for each group, the following application was made in the formula for seventh grade boys:

$$\text{C. R.} = \frac{p - q}{\sigma_p}$$

$$2.000 = \frac{p - q}{.054}$$

$$p - q = .1080$$

$$= 10.8\%$$

By similar applications for each of the other groups, the significant

<sup>3</sup>Ibid., p. 240.

<sup>4</sup>Lindquist, loc. cit.

differences in percent were derived as given:

Seventh grade boys	10.8
Seventh grade girls	11.0
Eighth grade boys	11.6
Eighth grade girls	11.6
Ninth grade boys	11.4
Ninth grade girls	10.4

In making the interpretations from the data tabulated from the questionnaire any difference between the total percent of Yes answers and the total percent of No answers that exceeded the above significant differences in the respective groups was considered statistically significant. Any percent difference that was less than the above significant differences was considered as due to chance only, and was not, therefore valid in this study.

# EQUIVALENCE OF THE PEARSON AND AYES FORMULAS OF CORRELATION

GEORGE C. KYTE  
University of California  
Berkeley, California

IN THE FIRST volume of the Journal of Educational Research, Leonard P. Ayres published his shorter method for computing the coefficient of correlation.<sup>1</sup> It is the method "which gives the sums of the products and the sums of the squares of the deviations directly from the squares of the original numbers." This method "avoids the necessity for taking into account the plus and minus signs of deviations."

Ayres' formula was based upon his stated principle: "This method is based upon considering every number in a series as being equal to the average of the series plus a plus or minus deviation from the average." Letting

Sub. = the first, or subject series

Rel. = the second, or relative series

Tot. = the total of the series, the sums of the items

Av. = the average of the series,

he formulated the statement of his method correlation as follows:

Sum of products of subject and relative items  
minus

av. of subject items  $\times$  tot. of relative items

---

$$\sqrt{\frac{\text{Sum of squares of sub. items} - \text{av.} \times \text{total of these items}}{\text{av.} \times \text{total of these items}}} \times \sqrt{\frac{\text{Sum of squares of rel. items} - \text{av.} \times \text{total of these items}}{\text{av.} \times \text{total of these items}}}$$

By the substitution of S for the subject items, R for the relative items, and N for the total number of items in each series, the Ayres formula<sup>2</sup> becomes:

<sup>1</sup> Leonard P. Ayres. "A Shorter Method for Computing the Coefficient of Correlation." Journal of Educational Research, I (March, 1920) pp. 216-221.

<sup>2</sup> Ibid. pp. 219-220.

$$r = \frac{\Sigma(S \times R) - \frac{\Sigma S \times \Sigma R}{N}}{\sqrt{\left[ \Sigma S^2 - \frac{(\Sigma S)^2}{N} \right] \times \left[ \Sigma R^2 - \frac{(\Sigma R)^2}{N} \right]}}$$

Having thus arrived at the formula, Ayres proved its equivalence to the Pearson formula by the application of empirical methods. He computed the coefficients of correlation for related sets of numbers whose distributions approximated normal probable curves, using both formulas. Obtaining the same value by each method, he thus established the equivalence of the two formulas. The similarity of the forms of the two formulas is obvious when the Pearson formula is written in the long form as follows:

$$r = \frac{\Sigma(x \times y) - \frac{\Sigma x \times \Sigma y}{N}}{\sqrt{\left[ \Sigma x^2 - \frac{(\Sigma x)^2}{N} \right] \left[ \Sigma y^2 - \frac{(\Sigma y)^2}{N} \right]}}$$

in which  $x$  stands for the deviations from the average of the first series and  $y$  stands for the deviations from the average of the second series.

The equivalence of the Pearson and Ayres formulas can be established algebraically. This proof is included below.

1. Since  $x = S - \text{av. } x$   
therefore  $x^2 = S^2 - 2S \times \text{av. } x + \text{av. } x^2$   
and  $\Sigma(x^2) = \Sigma(S^2) - 2\Sigma(S) \times \text{av. } x + N \times \text{av. } x^2$

$$\text{But } \text{av. } x = \frac{\Sigma(S)}{N}$$

$$\text{therefore } \Sigma(x^2) = \Sigma(S^2) - \frac{2(\Sigma S)^2}{N} + \frac{(\Sigma S)^2}{N}$$

$$\text{and } \Sigma(x^2) = \Sigma(S^2) - \frac{(\Sigma S)^2}{N}$$

2. Similarly  $\Sigma(y^2) = \Sigma(R^2) - \frac{(\Sigma R)^2}{N}$

3. Also  $x \times y = (S - \text{av. } x)(R - \text{av. } y)$

$$\text{and } x \times y = \left[ S - \frac{\Sigma(S)}{N} \right] \left[ R - \frac{\Sigma(R)}{N} \right]$$

$$\text{and } x \times y = RS + \frac{(\Sigma S)(\Sigma R)}{N^2} - R \left[ \frac{\Sigma(S)}{N} \right] - S \left[ \frac{\Sigma(R)}{N} \right]$$

$$\text{therefore } \Sigma(x \times y) = \Sigma(RS) + \frac{\Sigma S \times \Sigma R}{N} - \frac{\Sigma S \times \Sigma R}{N} - \frac{\Sigma S \times \Sigma R}{N}$$

$$\text{and } \Sigma(x \times y) = \Sigma(RS) - \frac{\Sigma S \times \Sigma R}{N}$$

Having derived the various values for the respective parts of the Pearson formula in terms of the S and R values of the Ayres formula, the writer has thus established mathematically the equivalence of the two formulas.

## BOOK REVIEWS

L. Joseph Lins, University of Wisconsin; Valworth Plum, University of Minnesota;  
John Schmid, Michigan State College

Frederiksen, Norman; Schrader, W. B. Adjustment to College.  
(Princeton, New Jersey: Education Testing Service, 1951), pp. xvii + 504.\*

This is a report of the study of 10,000 veteran and non-veteran students in sixteen American colleges undertaken in the years immediately following World War II. The study was supported by funds granted by the Carnegie Corporation of New York.

The report may be divided into two major sections. The first, based upon measures of academic ability and achievement, aims to compare veteran and non-veteran students with respect to academic success. The measure of success used is achievement relative to ability, rather than "absolute" achievement. The investigators found that veteran students were slightly more successful in general than non-veteran students.

The second portion of the study is based upon the responses of veteran and non-veteran students to a 46 item questionnaire. The questionnaire responses provided the basic data for separate chapters on characteristics of veteran students, their age, and general background, factors relating to motivation, their worries, how they spend their time, the GI bill, and their reactions to college.

While there is a wealth of data in these pages there are few surprising conclusions. In summarizing the differences between veterans and non-veterans, the authors report that "the similarities far outweigh the differences". The most clearly established differences have to do with age and maturity. There are, of course, many other indications of slight differences. These will be of interest to anyone who is making a special study of motivation, worry, or any of the other particular problems dealt with.

The report indicates that a great deal of careful and thoughtful work was done in planning the study and in analyzing the data. Special adaptations of standard statistical procedures, designed to fit the needs of this study and the characteristics of the data available, are discussed in the appendix. The broad scope of this study and the numerous complications involved in the use of existing records of test scores and grades in a variety of institutions must have presented serious problems to the investigators. The careful attention to detail, their ingenuity in overcoming obstacles, and their persistent effort to bring this study to a successful conclusion deserve commendation.

Robert L. Ebel

State University of Iowa

\*Because of limited supply, it is doubtful that copies of this book are available for distribution.

Jordan, A. M. Measurement in Education. (New York: McGraw-Hill Book Company, Inc., 1953) pp. xi - 533, \$5.25.

The teacher needs a continuous supply of evidence pertaining to the probable progress of pupils in all areas of living in order to provide adequate guidance to youth in their growth and development. It would appear, then, that the selection of adequate instruments of an evaluative character should assist teachers to plan wisely teaching-learning situations as well as the appraisal of learning outcomes.

In presenting his concept of measurement, the author states: "One fundamental concept, frequently illustrated, is that a test score is merely a sample of an individual's performance." (p. vii). "It is the purpose of measurement in education to furnish instruments for measuring more precisely the outcomes of education, to the end that the evaluation of them may not be dependent upon insufficient and uncertain evidence." (p. 4).

The book is divided into four areas: Part I, Problems of Measurement, which embraces discussions centered on characteristics of measuring instruments, the construction of achievement tests, the testing program; and measurement of reading, spelling, handwriting, language and literature, the social sciences, foreign languages, mathematics, science, business education, fine arts and manual arts, health, and physical education; Part II, Measurement of Intelligence; Part III, Personality Inventories; and Part IV, Statistical Methods.

The author deals directly with the problems of measurement as they relate to properly defined goals and the objectives of the educative process on both the elementary and secondary-school levels.

The book is well documented and includes a critical approach in discussing tests. Each chapter is concluded with a brief summary, practical questions, exercises, and a bibliography.

Teachers, supervisors, and administrators seeking information as to specific up-to-date tests related to subject matter, interest, attitude, and personality traits will find this text a welcome addition.

B. Everard Blanchard

Plymouth State Teachers College  
New Hampshire

Stone, Calvin P., Editor, Annual Review of Psychology, Vol. 4, (Stanford, California: Annual Reviews, Inc., 1953) pp. 462, \$6.00.

The 1953 edition of Annual Review of Psychology is the fourth in a series which began in 1950. The volumes are designed mainly for psychologists who are engaged in teaching and research and whose background knowledge is already well established. They are of value to educators because they compile, evaluate, and interpret the leading literature of nineteen psychological areas within the limits of a single volume. Many of these areas are of direct interest to educators.

This edition covers the same topics which were summarized in the preceding, 1952 edition, with two exceptions. A new section on special

disabilities is summarized by Katherine Cobb of Ridgewood, N. J., who was formerly with the Laboratory of Human Development, Harvard University. A further change involved establishing two separate sections on Physiological Psychology and Comparative Psychology which were formerly handled as a single section with joint authorship.

Of greatest direct interest to educators are the sections dealing with educational psychology, child psychology, individual differences, and counseling.

The section of Educational Psychology is summarized by H. D. Carter, University of California, Berkley, California. It considers such topics as readiness for learning, school learning, sociological correlates of learning, pupil personnel, instructional methods, research methods, evaluation, measurement, guidance and teacher personnel. It can be noted from this list of topics that the psychological aspects of almost every phase of education are touched upon. Amid this plethora of possible approaches to the psychology of learning, the author sees a need for isolation of really crucial problems and persistent research of a more penetrating nature than the fragmentary and isolated research which has been done.

The section on Child Psychology by D. B. Harris of the Institute of Child Welfare, University of Minnesota, is comprehensive in reviewing the literature, but is lacking in critical analysis of major trends or shortcomings of current research. The author has subdivided his section into topics covering infant behavior, perception, physical growth, intellectual development, communication skills, socialization, development of social behavior and development and measurement of attitudes, interest and personality in children.

The literature on Individual Differences is summarized in a brief section by Anne Anastasi of Fordham University. She considers developments in testing and evaluation as related to individual differences. She also reviews the applications of factor analysis procedures to the study of individual differences and the inherited and environmental conditions which account for them. The section concludes with discussion of extreme deviations and group differences.

The literature on Counseling, Therapy and Diagnosis is summarized by E. G. Williamson of the University of Minnesota. The author notes several trends in research in counseling. There appears to be a reduction in the number of studies on the processes and evaluation of outcomes of counseling and an increase in attempts to systematically formulate concepts of counseling in their practical aspects or in totality.

The format of this volume remains the same as the three preceding annual additions. Its utility would be increased if each contributor followed a more uniform plan in writing. Illustrative of this deficiency is the observation that nearly all present an introductory overview, but very few summarize trends and future needs for research. Users of the book are aided by a serial listing of literature which appears at the close of each section and by separate author and subject indices at the back of the book.

Professional educators should find this a ready reference to recent research in a variety of related psychological areas.

Lawrence P. Blum

Brief Commentaries

Garrett, Henry E. Statistics in Psychology and Education (New York: Longmans, Green and Co., 1953), pp. xii + 460, \$5.00.

This is the fourth edition since the original was published in 1926. The chief changes are a slight shortening in the length of the book, the inclusion of a new chapter dealing with analysis of variance, and a complete rewriting of the chapters on reliability and inference.

The book has undergone sectional reorganization being divided now into three major areas: (1) descriptive statistics, (2) prediction and inference, and (3) special topics (correlation and test construction).

-L. J. L.

Gruenberger, Fred. Computing Manual (Madison: The University of Wisconsin Press, 1952), pp. 123, \$2.00.

With greater and greater emphasis upon more rapid and accurate means of handling larger computing problems, individuals in their research work are turning to advanced techniques of machine calculation. This book is intended as a guide to semi-experienced computers and as a text and laboratory manual in the training of machine computers. Of primary concern is punched card computation through use of IBM (International Business Machines).

The book is divided into three sections: (1) Computing Notes, (2) Laboratory Work in Punched Card Computing, and (3) IBM Terminology. The first section treats questionnaire methodology, digitizing, cycling and selector control, bivariate tables and differencing, chi-square analysis, random sampling, miscellaneous operating hints, electronic calculation, and wiring diagrams. Problems for laboratory use are set up in the second section.

-L. J. L.

Tuckman, Jacob; Lorge, Irving. Retirement and the Industrial Worker (New York: Columbia University, 1953), pp. xvi + 105, \$2.75.

The research reported based upon study of the industrial worker gives important leads likewise to the problems confronting the educator prior to, at the time of, and during retirement. The study indicates that industrial workers approaching retirement age have a deep resistance toward retirement for reasons that stem from cultural, economic, psychological, and social factors.

-L. J. L.

Umstattd, J. G. Secondary School Teaching (New York: Ginn and Co., 1953), pp. xii + 488, \$4.50.

This, the third edition of Secondary School Teaching, a general methods book at the secondary level, stresses the unitary approach in teaching even more than did the former editions. To this end, a new chapter "The Co-operatively Planned Learning Process" has been added.

The book is brought up to date in statistics and bibliographies as well as in problems of contemporary life. The strength of the chapters on radio and visual aids is increased by the inclusion of television.

-L. J. L.

Wittich, Walter Arno; Schuller, Charles Francis. Audio-Visual Materials—Their Nature and Use (New York: Harper and Bros., 1953), pp. xx + 564, \$6.00.

How can I as a teacher with relatively limited educational resources create a teaching situation such that the students under my direction can more fully understand the world of today and can become well fitted to take their place in that world?

Through the medium of audio-visual materials, the authors point out how the world can be brought to the classroom. Emphasis is placed upon wise selection and effective use of audio-visual teaching aids. Covered are television, recordings, radio, motion and still picture projection, globes and maps, tackboard, graphics, three-dimensional materials and models, flat pictures, and the chalkboard.

Illustrations and pictures are chosen well and presented in abundance. The book is enhanced by suggested activities, a well-rounded bibliography, and source lists. The book reflects the wide experience and pioneering influence of its authors.

L. J. L.

## Research News and Communications

Lawrence P. Blum, University of Wisconsin Extension, Milwaukee; Jacob O. Bach,  
Southern Illinois University; Lester Sands, Santa Barbara College

### Research in field of spelling being conducted in the New York City Public Schools

An extensive experimental study in the field of language arts concerned specifically with spelling is being conducted cooperatively by the Division of Elementary Schools, the Division of Junior High Schools, and the Bureau of Curriculum Research.

Two years ago experimentation was begun with the introduction of Spelling Lists, A, B, and C. List C, the basic list from which the other two lists have been derived, was adopted from an unpublished list of words prepared by Gertrude Hildreth (Brooklyn College) and Ethel I. Salisbury (University of California at Los Angeles). In the Hildreth-Salisbury Alphabetical List each word was assigned a learning level number from 1 to 40 depending on the number of times it was written by children as reported in Henry D. Rinsland's study entitled "A Basic Vocabulary for Elementary School Children."

The evaluation of the program consists of four parts:

1. Study and recommendations by supervisors and teachers.
2. A survey of the writing of pupils.
3. A testing program.
4. Observation of teaching procedures and pupils' work.

Although this study will not be completed until 1954, the results to date suggest the following conclusions:

1. Supervisors and teachers concur in the opinion that the spelling lists are valuable instructional material for children and the accompanying manual includes valuable teaching suggestions.
2. A survey of the writing of 2700 children from grades 3 through 9 in the experimental schools indicates recognizable growth, appropriate to the children concerned, in their ability to spell words accurately in content.
3. A testing program showed improvement at all grade levels with the gains being greater at lower grade levels. This is in accord with other findings that growth in language skills appears greatest at age levels 8-10.
4. A program of observation by classroom teachers and language arts coordinators shows improvement in the participating schools in teaching procedures and in childrens' attitudes toward spelling.

Further information concerning this study can be secured from William H. Bristow, Director, Bureau of Curriculum Research, New York Public Schools, 130 W. 55th Street, New York 19, New York.

Educational Testing Service research fellowships again available.

The Educational Testing Service is offering for 1954-55 its seventh series of research fellowships in psychometrics leading to the Ph. D. degree at Princeton University. Open to men who are acceptable to the Graduate School of the University, the two fellowships each carry a stipend of \$2,500 a year and are normally renewable.

Fellows will be engaged in part-time research in the general area of psychological measurement at the offices of the Educational Testing Service and will, in addition, carry a normal program of studies in the Graduate School. Competence in mathematics and psychology is a pre-requisite for obtaining these fellowships. The closing date for completing applications is January 15, 1954. Information and application blanks will be available about November 1st and may be obtained from: Director of Psychometric Fellowship Program, Educational Testing Service, 20 Nassau Street, Princeton, New Jersey.

# JOURNAL of EDUCATIONAL RESEARCH

---

Volume XLVII

December, 1953

Number 4

---

## ITEM ANALYSIS AND DISCRIMINATIVE VALUE OF SELECTED WECHSLER-BELLEVUE SUBTESTS

EDMUND MECH\*  
Indiana University

IT IS PROBABLY THE EXPERIENCE of every psychometric examiner with most testing scales that some items which are supposed to be "easy" in view of their placement in the beginning of the scale are failed, while some of the more difficult ones placed at the end of the scale are passed. The occasional occurrence of such inconsistencies does not disturb the examiner and is attributed to the idiosyncrasies of the testee and to the peculiarities of his mental development, whether in the normal or abnormal range. This is also to be expected in view of the fact that final placement of items in order of difficulty is usually based on numerical summaries and statistical data of large standardization groups which mask individual patterns. However, when such inconsistencies occur with some degree of regularity and consistency, the correctness of the order of items is to be questioned, at least for the sample of the population involved.

It is observed that especially in such subtests as Information, Picture Completion, Comprehension, and Similarities, that some of the items which appear in the first part of the respective subtest are more difficult than the subjects and are failed with greater frequency than those appearing in the middle or last parts of these tests. It would seem then, that quantitative proof and substantiation of these empirical notions appears desirable.

### Problem

Wechsler's last edition of the Measurement of Adult Intelligence (3), shows an implicit recognition of the need for an analysis of item dif-

\*Institute of Educational Research.

ficulty. In fact, Wechsler adopts modifications in the order of presentations of the Information subtest questions based upon unpublished data summarized and communicated by Altus. Therefore, the problem in the main has two facets: (1) to determine the extent of difficulty of each item on the Wechsler-Bellevue (Form I) Intelligence Test, and (2) to obtain a relative index of discrimination for individual items. The bi-serial coefficient technique as applied to a continuous dichotomy in conjunction with a known criterion will be further elaborated upon later in the investigation. In pointing out that the study will be limited to a high school group on the senior level, an interesting notion is introduced, in that Wechsler's original standardization group only included 23 cases on the twelfth grade level, and a mere 100 cases on the entire four-year high school level. Although realizing the difficulties encountered in standardization, it is the opinion of the investigator that this is by no means an adequate measure of our secondary school students' performance on the Bellevue scale.

### Relative Importance of Problem

An intra-test analysis of item difficulty would accomplish two major services in a clinical situation, or in any counseling situation. In the first place, a more correct and statistically justified arrangement of items would make the test a tool of increased efficiency. It should be noted here that in several of the Wechsler subtests that the examiner is required to stop seeking responses after four or five consecutive failures on a subtest. It is necessary then to have the order of items arranged in a statistically justified manner in order to permit the clinician to obtain the optimum number of responses from a subject. Needless to say that with such an arrangement that time would be saved and further useless questioning obviated. Secondly, but no lesser in importance, it might be less discouraging to some sensitive testees who have to experience failures before reaching the actual limit of their capacities.

### Methods

According to Richardson and Stalnaker (2) when items are scored "1" if correct and "0" if incorrect, the assumption of normality in the distribution of responses to any given item is not warranted. The formula below gives a point bi-serial correlation coefficient which does not assume continuity in the distribution of single test items, and was applied in the ensuing analysis, in which

$$r_{bis} = \frac{M_p - M_q}{6} \cdot \sqrt{pq}$$

where  $M_p$  the mean of all correct responses and  $M_q$  the mean of all wrong responses;  $p$  is equal to the proportion of correct responses and  $q$  equals the proportion of wrong responses to any given item. The standard deviation ( $\sigma$ ) is calculated for the entire distribution.

### Description of the Sample

The subjects used in the investigation are 110 senior students from the Bloomington, Indiana, High School. They were administered the Wechsler-Bellevue, Adolescent and Adult Intelligence Test, Form I, at the Indiana University Psychological Clinic during the school year of 1948-1949.

Each subject voluntarily requested to take the Bellevue test, the results being used in aiding the school authorities in counseling and guiding the students. The mean I.Q. is 109.36, which according to the Wechsler classification, falls at the upper limits of the average group (91-109). Therefore the population mean can probably be classified as "high average." The standard deviation of the distribution is calculated to be 10.45. It would seem valid to assume that motivation was relatively good during the testing situation for each individual, due to the fact that testing appointments were made on a volunteer basis.

### Analysis of the Data

An analysis of the difficulty of the items in seven subtests of the Wechsler-Bellevue was undertaken. The Information, Comprehension, Similarities, and Arithmetic were included in the Verbal section. The Memory for Digits, forwards and backwards, test did not lend itself to a difficulty analysis, since the idea of difficulty depends on the number of digits repeated rather than differences in the difficulty of individual items. Three performance subtests seemed suitable for analysis: the Picture Completion, 15 items; Picture Arrangement, 6 items; and the Kohs Block Designs, which includes seven items. The Object Assembly Test is not included at this point, since there are only three items which comprise the test.

An inspection of the data in Table II shows the frequency of successful responses to each item as given by 110 high school seniors to the Information subtest of the Bellevue scale. The items have been given a rank order number corresponding to the degree of difficulty with which it presented each subject. Obviously the greater the number of subjects passing each item the easier the item is, thus the rank order of items 1, 2, 3, etc., is assigned from the easiest to the most difficult.

Comparing the observed number of passes with Wechsler's original

TABLE I

DISTRIBUTION OF 110 HIGH SCHOOL SENIOR SCORES ON  
THE WECHSLER BELLEVUE INTELLIGENCE SCALE, FORM 1

Wechsler I.Q.	Frequency	Wechsler I.Q.	Frequency
138-142	3	108-112	28
133-137	3	103-107	13
128-132	4	98-102	13
123-127	7	93- 97	12
118-122	9	88- 92	7
113-117	11		
<b>N= 110</b>			

TABLE II

SUCCESSFUL RESPONSES AND RANK ORDER ON INFORMATION SUBTEST MADE BY 110 HIGH SCHOOL SENIORS IN COMPARISON WITH WECHSLER'S ORIGINAL PRESENTATION

Wechsler Question Number	Number of Passes	Rank Order
1	110	1
2	108	2
3	100	3
4	92	8.5
5	80	14
6	96	5
7	76	15
8	72	16
9	84	12.5
10	84	12.5
11	96	5
12	40	18.5
13	96	5
14	86	11
15	30	20
16	93	7
17	14	24
18	92	8.5
19	90	10
20	40	18.5
21	44	17
22	19	22
23	21	21
24	5	25
25	17	23

order which is presented in the left margin of Table II, it is immediately apparent that some items, which according to the Bellevue scale should be gradually more difficult, are versa. For example, Wechsler's item 5, "How many pints make a quart?" is ranked fourteenth in the present investigation. A popular response among the wrong answers was that four pints make a quart. It is not within the scope of the present investigation to venture a theory as to why approximately 28 per cent missed the item, whether it be due to temporary inefficiency or other factors. Item 17, "Who discovered the North Pole," proved to be exceptionally difficult for the population involved, ranking twenty-fourth on the scale due to only 13 percent passing the item. The popular reply to this item was "Admiral Byrd." It can also be seen from Table II that items 4, 8, 9, 12, 15, 16, 18, 19, and 21 are not in accordance with Wechsler's order, and there is much variation.

Up to this point the only things known are a quantitative number of subjects passing each item, assigning the item a proper rank order and comparing with Wechsler's presentation order. This information is valuable in that it will enable the investigator to suggest a revised order for the population involved, but at the most can only give an approximate order for presentation.

It is the opinion of the investigator that it is also pertinent to attempt a measurement of "discrimination" between items. The rationale is thus, an item that can be passed equally often by individuals of all-levels of ability obviously has no discriminating power. Thus items that are entirely too easy or entirely too difficult are not diagnostic of differences of ability. They, also, are not very diagnostic.

Theoretically, every test item can be passed by more and more individuals as the level of ability increases from group to group. Thus in the treatment of the item analysis or discrimination the investigator proposes to apply the bi-serial method of correlation to the data, to find whether success or failure upon a given item is correlated with total score in the test, which happens to be the criterion in this case. If those who achieve high scores in the criterion get an item right more often than those who make low scores, the item will be positively correlated with the criterion. Such an item is a better measure of a criterion than one which correlated zero or negatively with criterion scores.

Table III presents the point bi-serial correlations made for each item on the Information subtest. It is readily apparent from Table III that the most discriminating items are numbers 5, 7, 8, 9, 10, 12, and 17.

Table IV illustrates the relative order in which the Comprehension subtest items were ranked on the basis of successful responses. The successful responses on this particular subtest can be separated into two groups, a complete response to each item is credited with a score of 2.

TABLE III  
 POINT BI-SERIAL CORRELATIONS FOR EACH  
 ITEM ON THE INFORMATION SUBTEST  
 WITH BELLEVUE FULL SCALE

Item	Point Bi-serial r
1	...
2	.184
3	.221
4	.252
5	.457
6	.236
7	.303
8	.423
9	.382
10	.304
11	.085
12	.449
13	.229
14	.276
15	.021
16	.237
17	.612
18	.096
19	.103
20	.226
21	.297
22	.217
23	.295
24	.017
25	.062

TABLE IV

NUMBER OF SUCCESSFUL RESPONSES ON THE  
COMPREHENSION SUBTEST

Item	Successful Responses	Rank Order
1	108	1.5
2	106	3.5
3	108	1.5
4	106	3.5
5	103	5
6	97	6.5
7	69	9
8	97	6.5
9	45	10
10	74	8

TABLE V

DISTRIBUTION OF BI-SERIAL COEFFICIENTS  
ON COMPREHENSION SUBTEST

Item	Bi-Serial Coefficient	
	1 credit	2 credits
1		
2	.032	.107
3		
4	.137	.274
5	.102	.148
6	.179	.472
7	.237	.495
8	.168	.752
9	.171	.000
10	.047	.187

While an incomplete response is credited with a score of 1, a failed response is credited with a score of 0. For the purpose of being able to compute order of difficulty without confusion it was considered expedient to group the item as a "pass" if scored either 1 credit or 2 credits. It is clear that much variation exists in the latter part of the subtest, especially among the last four items. Only 40 per cent of the population were successful on item 9, "Why does the state require people to get a license in order to be married?" and it ranks a solid number 10 in difficulty, while Wechsler's original item 10, "Why are people who are deaf usually unable to talk?" was passed by 70 per cent of the group. Item 7, "If you were lost in a forest in the daytime, how would you go about finding your way out?" was passed by only 63 per cent of the population and is ranked as number 9.

Table V illustrates distribution of the bi-serial coefficients on the Comprehension subtest.

As was mentioned before, the scoring to each item on the Comprehension subtest can be either 0, 1, or 2 credits. It is the opinion of the investigator that in this case it is not valid to simply correlate the number of subjects who "pass" each item, with the criterion, but to consider the question of "How well did each item correlate with total score when the subjects who scored as 1 credit were measured with it as opposed to those who scored 2 credits?"

There is not an overabundance of rationale advanced for each subtest, and Wechsler does not offer very much in his revised edition. Rapaport (1) attempts to present a psychological rationale for the Comprehension subtest, in that the function underlying the Comprehension subtest is related to the function of "judgment." Rapaport also contends that the Comprehension items, like situations calling for judgment, require more than merely a delay of impulse and an enumeration of possibilities. In question 9, "Why does the state require people to get a license in order to be married?" references to license revenue, legality of children, health examination, age requirements, prevention of venereal disease, do not improve the response; they rather call for an appropriate selection. Probably what Rapaport is attempting to reach in his rationale is that the Comprehension test attempts to measure a level of "social awareness." Probably the best example of this would be item 8, "Why are laws necessary?" which had a  $r_{pbis}$  of .171 with 1 credit and a bi-serial coefficient of .000 at the 2 credit level, would mean that those who scored higher on the criterion tended to score 1 credit oftener than those who did not, but the  $r_{pbis}$  is sufficiently small as to be of negligible value, and of no use as a discriminating item. In other words, it did not differentiate between groups. The most significant bi-serials were found to be with items 6, 7, and 8.

Table VI presents the successful responses on the Similarities subtest and relative rank order.

TABLE VI

SUCCESSFUL RESPONSES ON THE SIMILARITIES  
SUBTEST AND RELATIVE RANK ORDER

Item	Passes	Rank Order
1	110	1
2	108	3
3	108	3
4	101	5
5	108	3
6	53	10
7	71	9
8	93	7
9	99	6
10	43	11
11	39	12
12	87	8

TABLE VII

BI-SERIAL COEFFICIENTS FOR EACH ITEM  
ON THE SIMILARITIES SUBTEST

Item	Bi-Serial Coefficient	
	1 Credit	2 Credits
1		
2		
3		
4		
5		
6	.215	.297
7	.217	.374
8	.131	.119
9	.149	.103
10	.074	.004
11	.137	.014
12	.044	.573

TABLE VIII

SUCCESSFUL RESPONSES TO INDIVIDUAL  
ITEMS ON THE ARITHMETIC SUBTEST

Item	Number of Passes	Rank Order
1	110	1.5
2	110	1.5
3	108	3
4	102	4
5	91	6
6	101	5
7	71	7
8	45	8
9	13	9
10	11	10

TABLE IX

DISTRIBUTION OF BI-SERIAL COEFFIC-  
IENTS ON ARITHMETIC SUBTEST

Item	Bi-serial Coefficient
1	...
2	...
3	.134
4	.201
5	.210
6	.236
7	.274
8	.384
9	.302
10	.114

Wechsler considers the Similarities subtest as among the best of the entire battery. It is possible to get here an obvious difference as to a level of thinking in relation to an individual's ability to discriminate between essential and superficial likenesses. Thus when a subject says an orange and banana are alike because "you can eat them" and a bicycle and wagon "because they have wheels," he receives a credit of 1, whereas the responses, "both are fruits," and "means of conveyance," are scored 2. This qualitative difference in response is of value because it is often suggestive of the level of the subject's intellectual functioning.

Table VI shows the results of the responses on the Similarities subtest, much variation again being found for the population involved. Item 6 on Wechsler's list ranks tenth. Item 7 is ranked as ninth, while the largest discrepancy found was that Wechsler's supposedly most difficult item 12 ranked only eighth in difficulty.

Table VII shows the bi-serial coefficients for items in the Similarities subtest.

There was no need to apply any technique to the first five items because no valid bi-serial could be computed, as can be seen from Table VI. There are not enough "wrong" responses. The best discriminating items are at the 2 credit level. Item 2 gives a  $r_{pbis}$  of .573 and item 7 gives a correlation of .374. Although item 12 is ranked only eighth in difficulty among the population sampled it proved to be the best discriminator. Subjects who did receive 2 credits on this item tended to receive higher total I.Q.'s.

Table VIII illustrates successful responses to individual items on the Arithmetic subtest.

The original Wechsler order on the Arithmetical Reasoning Test seems to be compatible with the rank order of the population involved. A suggested change would be valid in reversing questions 5 and 6. The rank order can be seen in Table VIII.

Question number 8, "A man bought a second-hand car for  $\frac{2}{3}$  of what it cost new. He paid \$300 for it. How much did it cost new?" seems to differentiate significantly between those that score higher on the total test.

#### Performance Subtest Analysis

The Block Design Subtest is, according to Wechsler and others, the best single performance item. It is actually modified from Kohs and is similar to that as employed by him in his original standardization. Wechsler shows much enthusiasm for it and sums the rationale up by stating that the block designs, whatever the mechanism involved, is good for showing up early as well as late disturbances in the higher perceptual processes, such as organic brain disease. Table X shows

TABLE X

## SUCCESSFUL RESPONSES ON BLOCK DESIGN SUBTEST

Item	Passes	Rank Order
1	110	1.5
2	110	1.5
3	105	3
4	107	4
5	74	5
6	59	7
7	67	6

TABLE XI

POINT BI-SERIAL COEFFICIENTS FOR  
BLOCK DESIGN SUBTEST

Item	Bi-serial Coefficient
1	...
2	...
3	.031
4	.177
5	.310
6	.598
7	.351

TABLE XII  
SUCCESSFUL RESPONSES TO EACH ITEM ON THE  
PICTURE COMPLETION SUBTEST

Item	Number of Passes	Rank Order
1	109	1
2	108	2.5
3	108	2.5
4	78	11
5	71	12
6	106	5
7	93	9
8	105	6
9	107	4
10	103	7
11	51	14
12	83	10
13	41	15
14	59	13
15	98	8

the number of successful responses to the Block Designs Test.

From Table X, the only variation from original order was between items 6 and 7. This is a curious discrepancy because item 7 is the "big" design involving 16 cubes and item 6 involves only 9 cubes. Actually item 7 is the most complex, though from the view of visual organization is not the most difficult. It would seem that the amount of work required in item 7 is its most difficult asset, but in a homogeneous population as used in the present study, the overwhelming amount of work did not prove as difficult as it might have with a population composed of depressive patients. Item 6 requires diagonal orientation and has a time limit as all the others do. Possibly many could have completed it successfully if more time was available. It seems to prove more difficult to those subjects who cannot get the idea quickly. In any case, item 6 proved more difficult than item 7.

Table XI shows the bi-serial coefficients of each item. Item 6 which ranked seventh in difficulty is the best differentiating item in the Block Designs subtest. Item 7 which ranked sixth was next best with a .351 bi-serial coefficient. In toto, the Block Designs test produced three items out of five used that are successful in differentiating between those who receive higher total scores.

The Picture Completion subtest, which consists of 15 single pictures, is in short a test of one's ability to see what is missing from any particular picture. To know what is missing, the subject must first know what that picture represents. Wechsler maintains that it is particularly effective in picking out mental defectives and is relatively inadequate in discriminating between higher levels of intelligence. Let us say it has a "low ceiling," there being a tendency for scores to pile up at the upper end of the distribution. Table XII shows the distribution of responses on the Picture Completion subtest.

It is obvious from Table XIII that much variation in order of difficulty exists on the Picture Completion subtest. Item 4, as ranked by Wechsler, ranks eleventh and item 5 as ranked by Wechsler is placed twelfth. Item 15 on the Bellevue scale is ranked but eighth in difficulty in relation to the population involved. Item 11 is ranked fourteenth and item 8 is ranked but fourth. There is need for a revised order of presentation for this subtest.

Although the Picture Completion Test produced much variation in order of difficulty, it did not produce but two items that could discriminate significantly between those who received higher total scores and those that received lower ones. Item 4 with a bi-serial coefficient of .383 and item 5 with a coefficient of .422 were the best differentiating items. It seems that the Picture Completion subtest is not too adequate for discriminating on the higher intellectual levels as measured by the Bellevue scale.

The Picture Arrangement Test, which is the last one to be considered consists of a series of pictures which, when placed in the right sequence

**TABLE XIII**  
**POINT BI-SERIAL CORRELATIONS FOR**  
**PICTURE COMPLETION SUBTEST**

Item	Bi-serial Coefficient
1	...
2	.243
3	.154
4	.383
5	.422
6	.211
7	.239
8	.227
9	.079
10	.207
11	.094
12	.197
13	.043
14	.186
15	.091

December 1953)

MECH

257

TABLE XIV

## SUCCESSFUL RESPONSES ON PICTURE ARRANGEMENT SUBTEST

Item	1	2	3	4	5	6
Passes	110	108	103	87	61	55
Rank Order	1	2	3	4	5	6

TABLE XV

POINT BI-SERIAL COEFFICIENTS FOR PICTURE  
ARRANGEMENT TEST

Item	1	2	3	4	5	6
Bi-Serial Coefficient	...	...	.203	.334	.279	.307

TABLE XVI

## SUGGESTED ORDER OF ITEM PRESENTATION

Item	Information	Picture Completion	Similarities	Comprehension	Arithmetic	Block Design	Picture Arrangement
1	1	1	1	1.5	1.5	1.5	1
2	2	2.5	3	3.5	1.5	1.5	2
3	3	2.5	3	1.5	3	3	3
4	8.5	11	5	3.5	4	4	4
5	14	12	3	5	6	5	5
6	5	5	10	6.5	5	7	6
7	15	9	9	9	7	6	
8	16	6	7	6.5	8		
9	12.5	4	6	10	9		
10	12.5	7	11	8	10		
11	5	14	12				
12	18.5	10	8				
13	5	15					
14	11	13					
15	20	8					
16	7						
17	24						
18	8.5						
19	10						
20	18.5						
21	17						
22	22						
23	21						
24	25						
25	23						

tell a little story. The picture series are not unlike the short comic strips to be found in the daily papers. The pictures are presented to the subject in a disarranged order and he is asked to put them together in the right order so that they make a sensible story. The only conclusion that Wechsler can come to is that individuals who do fairly well on the Picture Arrangement seldom turn out to be mental defectives, even when they do badly on other tests. Table XIV shows the distribution of correct responses to each item.

No variation was found as to order of presentation in the Picture Arrangement subtest. Table XV shows the distribution of bi-serial coefficients for the Picture Arrangement test.

The Picture Arrangement subtest produced two significant differentiating items. The most significant was item 4 with a bi-serial  $r$  of .334. Item 6 followed with a bi-serial  $r$  of .307.

Table XVI shows a suggested revised order of presentation of the Wechsler-Bellevue Intelligence Scale as it pertains to students on the senior high school level. Much variation has been found among the various subtests as the data presented has shown.

### Summary and Conclusions

Experience during the administration of the Wechsler-Bellevue scales prompted a detailed quantitative analysis of the difficulty and discriminatory power of items based on the 110 records of apparently normal high school senior students. The analysis was confined to the following seven subtests which lent themselves to such statistical treatment: Information, Comprehension, Similarities, Arithmetic, Block Designs, Picture Completion, and Picture Arrangement, a total of 85 individual items in all. The process was to rank the items in their order of difficulty for the senior high school population, and compare the results with Wechsler's original order of presentation. The second step was to compute a point bi-serial coefficient of correlation for each item in order to obtain a measure of discrimination.

### Conclusions

On the basis of the results in the investigation, it seems valid to conclude the following points:

1. That a modified revision should be carried out in the original order of item presentation. The investigator would like to point out that this applies only to students in the twelfth grade.
2. It is felt that the findings may help speed up the administration of the test. The suggested changes in the original order of item presenta-

tion are summarized in Table XVI. A revised order would avoid a frequent and excessive sense of failure on the part of less gifted examinees.

3. The Verbal subtest items produced a greater percentage of discrimination, in general, between those who scored higher on the total test and those who made lower I.Q. scores. The two best subtests, on the basis of number of discriminating items in the Verbal section are Comprehension and Information.

4. The best single Performance subtest on the strength of point bi-serial coefficients, seems to be the Block Designs. The least best is the Picture Completion, which produced only two significant items out of a possible 15.

#### Implications of the Study

A more statistically justified test should help speed up the administration time and of equal importance, avoid a frequent and excessive sense of failure on the part of less gifted examinees. This seems especially pertinent because many educational institutions through the medium of a psychological clinic, are using the results of the Bellevue scales in counseling and guiding the graduating senior students.

The suggested order revision would aid both the clinician and client, as the purpose of an individual test is to give the client a much more lenient atmosphere and obtain from him an optimum performance. Items that are badly misplaced serve to destroy rapport, which is all important to a valid testing situation.

#### REFERENCES

1. Rapaport, David. Diagnostic Psychological Testing, Vol I (Chicago: Year Book Publishers, Inc., 1946).
2. Richardson, M. N., and Stalnaker, J. C. "Use of Bi-Serial r in Test Research," Journal of General Psychology, VIII (1933), 463-465.
3. Wechsler, David. The Measurement of Adult Intelligence, 3rd ed. (Baltimore: Williams and Wilkins Co., 1944), 254 pp.

## ADDENDA

The following are corrections and additions to Victor Mech's article which is published in this December, 1953, JOURNAL OF EDUCATIONAL RESEARCH, entitled "Item Analysis and Discriminative Value of Selected Wechsler-Bellevue Subtests." The insertions are underlined and quotation marks should have been used as indicated.

\* \* \* \*

The lines including "it is the experience of every psychometric examiner.... the correctness of the order of items is to be questioned, at least for the sample of the population involved" should be in quotes and preceded by, The need for item difficulty research is cogently pointed out in the following statements by Rabin, Davis and Sanderson (1946, p. 493):

The lines including "It is observed that especially in such subtests as Information.... substantiation of these empirical notions appears desirable" should be in quotes and preceded by, With specific reference to the Wechsler test, Rabin, Davis, and Sanderson (1946, p. 493) have observed that:

The following sentence should be inserted immediately prior to the section headed The Problem, It was the purpose of this paper to follow up the notions generated by Rabin, et al (1946) using a sample of senior high school pupils.

The lines including "Wechsler's last edition.... unpublished data communicated by Altus" should be in quotes and preceded by, More specifically, Rabin, Davis and Sanderson (1946, p. 493) state:

The sentences including "In pointing out that the study.... of our secondary school students' performance on the Bellevue scale." should be corrected to read as follows: "In pointing out that the present study was limited to senior high school pupils, it perhaps is of interest to note that for standardization purposes, Wechsler (1944, p. 114) reports sampling approximately 200 twelfth grade pupils between the ages of 15 and 16 years."

The lines including "an intratest analysis.... services in a clinical situation" should be in quotes and preceded by, Rabin, Davis and Sanderson (1946, p. 493) indicate the relative importance of the problem by stating:

The lines including "it might be less discouraging to.... reaching the actual limits of their capacities" should be in quotes and preceded by, Secondly, but no lesser in importance, Rabin et al (1946, p. 494) point out that:

Add to the references: Rabin, A. I., J. C. Davis, and M. H. Sanderson, "Item Difficulty of Some Wechsler-Bellevue Subtests," Journal of Applied Psychology, XXX (1946), pp. 493-500.

# TWO ADDITIONAL STUDIES IN THE DYNAMICS OF SCHOOL SOCIAL STRUCTURE OF CLASS- ROOM SEATING AND SCHOOL DANCES

E. L. GREENBLATT  
Manhattan Beach City Schools  
Manhattan Beach, California

## I. Classroom Seating in Relation to Mental Health and Sociometric Status

THE PURPOSE of this study was to determine to what extent class seating, on the basis of unlike mental health status, will effect growth in mental health as compared to the growth patterns of a comparable class, its organization unchanged as a result of mental health and sociometric testing.

During this period, the school year of 1949-50, an investigation was also made into the relationship of a child's degree of social participation and that child's mental health and sociometric status.

### Subjects

The children used in the study were those of two seventh grade classes, Center School, Manhattan Beach, California. A child's mental health status number 1 (see Table I) is the raw score of the child on the California Mental Health Analysis, Form A, administered October, 1949. The sociometric status number 1 (Table I) is that determined by a sociometric questionnaire of the same date. Comparison of the two classes, on the basis of achievement, age and ability, found them to be very similar in these areas. A detailed account of this comparison and the sociometric techniques employed may be found in the study, "Relationship of Mental Health and Social Status".<sup>1</sup>

### Materials and Procedures

The classes of the study were designated Class A and Class B. The children of Class A were seated on the basis of unlike mental health status; the children of high mental health status being seated with the children of low mental health status. In making this arrangement, factors of mental age, chronological age and achievement were taken into consideration so as not to place a child in an environment which might tend to frustrate his efforts to gain academic and social success.

<sup>1</sup>Journal of Educational Research, November, 1950.

TABLE I  
RELIABILITY OF THE MENTAL HEALTH AND SOCIOMETRIC TESTS

Device correlated	Class A		Class B		Total Sample
	Boys	Girls	Boys	Girls	
<b>Sociometric status No. 1 with</b>					
Sociometric status No. 2	.81	.76	.73	.93	.83
<b>Mental Health No. 1 with</b>					
Mental Health No. 2	.84	.87	.76	.68	.80

However, the discrepancies in mental health status were kept at a maximum.

Class B progressed through the school year unchanged as a result of information gained from sociometric and mental health testing.

The last week of school the children in both classes were given the California Mental Health Analysis, Form A, and the sociometric questionnaire. The results of these tests will be termed mental health status number 2 (see Table I) and sociometric status number 2.

### Results and Interpretation

Prior to examination of the results in terms of relative growth in mental health, it may be well to examine the reliability of the measuring devices used. In Table I the results of the mental health and sociometric tests number 1, October, 1949, were correlated with the results of the mental health and sociometric tests number 2, June, 1950. These correlations, with a sample of 65 children, were made on the basis of class, sex, and total sample. The standard error for the total sample was .125. The standard error for the sample of 16 boys in Class A was .25, .26 for the 17 boys in Class B, and .25 for the 16 girls in each class.

The correlations seem sufficiently high to assume reliability for the devices used. Regarding the sociometric tests it should be noted that the high degree of reliability does not assume that a child has chosen the same person or persons on the second test as on the first, but rather that a child will tend to a strong degree to maintain his rank on the questionnaire over a period of time.

Mental health and sociometric scores listed in Table II, and their gains or losses, are average scores of the indicated samples.

The results in Table II appear to indicate that there will be no appreciable difference in gain in mental health status as a result of seating children of differing mental health levels together as compared to the gain of children in a class in which seating arrangements are made on other bases.

It is interesting to note that not only did girls of both classes maintain their lead in average mental health score during the school year, but also increased the initial 7.9 advantage to 10.8 points over the boys.

In the study made earlier, an examination was made of the relationship of children choosing one another on the sociometric questionnaire. It was found that mutual choices tend, to a strong degree, to have like mental health levels, and that there was little or no relationship between the sociometric scores of mutual choices. An investigation of these findings on the basis of the second mental health and sociometric

TABLE II

CHANGES IN MENTAL HEALTH AND SOCIOMETRIC STATUS IN CLASSES A  
AND B AND IN THE TOTAL SAMPLE

	Girls	Class A Boys	Total	Girls	Class B Boys	Total	Total Sample Girls	Total Sample Boys
Mental Health No. 1	147.9	140.3	144.1	151.9	143.9	147.9	150.0	142.1
Mental Health No. 2	159.4	149.6	154.5	162.6	150.9	156.7	161.0	150.2
Gain	11.5	9.3	10.4	10.7	7.0	8.8	11.0	8.1
Sociometric No. 1	47.0	36.8	41.9	44.6	35.5	40.0	45.8	36.1
Sociometric No. 2	40.5	34.7	37.6	40.4	34.8	37.5	40.4	34.7
Loss	6.5	2.1	4.3	4.2	.7	2.5	5.4	1.4

tests was made. In the first questionnaire, 211 mutual choices were found to occur throughout the seven questions of the form. The correlation between the mental health scores of mutual choices for the entire questionnaire was .61. When confined to a question of purely social implication, with a sample of 40 cases, the correlation between mental health scores of mutual choices rose to .81.

It is interesting to note that of the initial mutual choices, only 50 remained. The correlation between the mental health scores of these, on the basis of the entire questionnaire and on the question of social implication was .32 and .48 respectively; the latter correlation resulting from a sample of 15 cases.

It therefore appears that the earlier conclusion, that children of like mental health levels will choose one another, must be modified to state that this may be true at the start of a school year but does not appear to retain its strength over a period of time.

The second finding mentioned of the earlier study, that there is little or no relationship between the sociometric scores of mutual choices, seems to be substantiated by the correlation of .11 based on the second sociometric tests. The 'r' in the first study of the sociometric scores of mutual choices was .12.

In correlating the mental health score with total sociometric score of individual children, the resulting 'r' of .37, with a sample of 65 children, was considered insignificant. The conclusion, that there was little or no relationship between the sociometric score of a child and that child's mental health status, is strengthened by the finding of a correlation of .18 between these factors based on the second series of tests. The standard error in both instances is .125.

## II. Participation in School Dances in Relation to Mental Health and Sociometric Status

To obtain a measure of social participation, three school dances were attended by the author and an assistant. The interval between dances was two weeks. They were confined to the seventh grade classes and attendance was not compulsory. The dances were planned so that there were an equal number of both girls' and boys' choices. Actual sampling was delayed until thirty minutes had elapsed to permit the children to overcome their initial anxieties. After this period, a child was assigned one point each time he was observed "sitting-out" a dance. In this way a measure of lack of participation was obtained.

### Results and Interpretation

It was decided to examine the sample of two bases, as follows: Children attending all three dances, two dances, one dance or no dances,

and on the basis of sex. This was done since it appears hazardous to stipulate reasons for non-attendance on the part of the children, and also to prevent differences in social customs between boys and girls from distorting the analysis.

The results as shown in Table III indicated that for those attending two dances and three dances there appears to be an inverse relationship between average mental health score and non-participation points. That is, the higher the mental health status the lower the point assessment for lack of participation. High scores on the Mental Health Analysis are indicative of a greater degree of adjustment on the part of the child. It should also be pointed out that points for non-participation are consistently higher for girls than boys. This may indicate that although the sponsors of the dances provided an equal number of girls' and boys' "choices", the opportunity did not dissuade the girls from the belief that it is the privilege of the male to request a dance companion.

The sample from which the above information was tabulated consisted of 31 cases: 16 boys and 15 girls. The standard errors were .25 and .26 respectively.

Examination of the results of the correlations for the most part seem to substantiate the belief that the higher an individual's status in mental health, the less he will be found in the role of a non-participant. The correlation between the total mental health score and non-participation points of girls, -.45, indicates this inverse relationship; the higher the mental health score the lower will be the total points for lack of participation. The correlations of the sub-titles of the Mental Health Analysis, Behavior Immaturity and Feelings of Inadequacy with non-participation points, -.94 and -.54, for girls, appears to strengthen the above statement.

The inverse relationship between total non-participation points and questions number one and seven on the sociometric questionnaire appears to exist, although to a considerably less significant degree than the correlation between total sociometric score and total non-participation points.

The correlations obtained for the above factors for the boys are not of sufficient strength to warrant analysis on the basis of the inverse relationships previously discussed. It may well be that the measures obtained for the boys are in a large degree influenced by many other factors (possibly his efforts to overcome timidity and to assume society's approved role of masculine aggressiveness). If this appears sound, then the measures for the girls may largely be of the degree of desirability.

#### Further Study

The problem of sampling in a large room with a constantly shifting arrangement of personnel has probably led to some error in this study.

TABLE III

**ANALYSIS OF SAMPLE ON THE BASES OF SEX, MENTAL HEALTH,  
NON-PARTICIPATION POINTS AND DANCE ATTENDANCE**

Dances	Number Attending		Average Mental Health Score		Average non-par- ticipation points	
	Girls	Boys	Girls	Boys	Girls	Boys
0 dances	1	2	148.0	127.0		
1 dance	9	5	146.1	142.1	2.77	.0
2 dances	8	11	140.7	149.7	7.00	3.18
3 dances	15	16	147.5	140.7	6.80	6.18

TABLE IV

CORRELATIONS OF TOTAL NON-PARTICIPATION POINTS OF CHILDREN ATTENDING  
 ALL THREE DANCES WITH TOTAL MENTAL HEALTH SCORE AND SUB-TITLES  
 AND TOTAL SOCIO METRIC SCORES AND QUESTION 1 AND 7  
 OF SOCIO METRIC QUESTIONNAIRE\*

---

Total Mental Health with Non-participation points	.45	-.04
Behavior Immaturity with Non-participation points	-.94	.29
Feeling of Inadequacy with Non-participation points	-.54	.20
Total Sociometric Score with Non-participation points	-.53	-.03
Question No. 1 with Non-participation points	-.10	-.10
Question No. 7 with Non-participation points	-.37	.22

---

\*Question 1 was, "Who would you like for your best friend?"  
 Question 7 was, "Who would you like to go to dances with?"

Many very timid children have a way of discovering hidden recesses in which to escape participation in the dances, though their presence indicates their desire to participate. Still others spend more time than necessary in rest rooms. Obviously, keeping an accurate check on these children was not possible. However, at no time were the children aware of the sampling process.

It is believed that further study in other areas of social participation, in which the control problem is less acute, is indicated.

### Conclusions

In relation to stimulation of growth in mental health as influenced by seating, the experiment indicates that:

1. Growth in mental health of children in a class in which organization is not based upon information gathered from sociometric and mental health tests is not inferior to the growth in mental health of a class in which the children are seated on the basis of unlike mental health. The growth in the two classes examined on the bases of average mental health growth and individual growth in mental health status show similar results. The small differences in favor of the control class are not significant.
2. Changes in sociometric status in both classes failed to reveal any significant differences between the classes as a result of the experiment.
3. Girls at this grade level score considerably higher in mental health status at the start of a school year and not only maintain that lead during that period, but also increase it to a small extent. This is equally true of the experimental and control groups.
4. The earlier finding, the children forming mutual choices will tend to a large extent to have a like mental health level, must be revised to state that this relationship does not seem to continue over an extended period of time, but rather is evident only at the beginning of the school year. The finding that there is little or no relationship between the sociometric scores of mutual choices and the sociometric and mental health scores of individual children, is, however, substantiated by this study.
5. The establishment of control groups on the basis of like mental health levels, like sociometric levels, and unlike sociometric levels, may yield valuable information in the field of mental health and social status of adolescent children.

In relation to participation in dances as related to mental health and social status, the study also shows that:

1. Girls with high mental health status and high sociometric status are found in the role of non-participant to a considerably lesser degree than girls having a low mental health and sociometric status.
2. The findings concerning boys of this study are inconclusive.

# PUPIL PREFERENCE FOR TITLES AND STORIES IN BASAL READERS FOR THE INTER- MEDIATE GRADES

MARGARET L. DRONEY, Lowell, Mass.,  
STELLA M. CUCCHIARA, Everett, Mass.,  
ALICE M. SCIPIONE, Framingham, Mass.\*

## Purpose of the Study

IT SEEMED worthwhile to examine the contents of intermediate grade basal readers to determine objectively children's interests in the titles and stories of such readers, and to provide children with an opportunity to state their choices for materials judged by publishers to be appealing to them.

More specifically, the purpose of this study is (1) to determine which story titles in basal readers appeal to children in grades IV, V and VI; (2) to compare the choices of boys and girls for story titles; (3) to determine whether children select a title because of title appeal or story appeal; (4) to compare sex differences in story and title appeal; (5) to compare the preferences of the upper and lower quartile for title appeal and story appeal.

## Method of Procedure

Making the Instruments. —The investigation analyzed the titles and stories in the basal readers published by D. C. Heath and Company; Macmillan Company; Scott, Foresman and Company; Silver, Burdett Company; and Ginn and Company. It consisted of two questionnaires with the first one devised to compare the titles within each basal reader in order to establish the ten most popular titles in each book. A separate list of titles was compiled from each book; and since titles in basal readers are generally arranged by specific classifications, the authors rearranged the order in which the titles appeared in the books. This procedure was considered necessary in order to eliminate as far as possible only a partial reading of a list of titles and to equate more nearly each title's chance of being selected. The five lists of titles were stapled together and one list presented to the children on each of five consecutive days. Each child was allowed time enough to read the entire list of titles and was free to ask the teacher for any assistance needed in the reading of the titles. The children were then asked to indicate with an "X" the five titles which most appealed to them. They were not asked

\* Public Schools of Lowell and Everett, Massachusetts; State Teachers College, Framingham, Massachusetts.

to indicate their degree of preference for any of the five titles selected.

A second survey was initiated to determine children's interests in titles as compared to children's interests in stories. A list for each grade, containing fifty titles—the ten most highly rated from the five basal readers used in the first survey—was arranged in random order so that all titles from one reader would not appear in succession. The second instrument consisted of two pages with twenty-five titles on each page and was administered in two sittings. After each title were placed the symbols "L", "I", and "D". The children placed a cross under the L if they liked the title, and an X under the D if they disliked the title. In order to determine whether a child's choice for a title was based on title or story appeal, it was necessary to include an additional section with the question, "Have you read the story?" The child placed an X under the "Yes" to indicate he had read the story, and X under the "No" if he had not read the story, and X under the "Don't know" if he could not remember having read the story.

Distribution of the Instrument. —The investigation was conducted in the fourth-, fifth-, and sixth-grade classrooms in the following Massachusetts communities: Everett, Wakefield, Lowell, Watertown, Belmont, and Winchester; and in Nashua, N. H. The first survey involved 1498 children and the second, 744 children, equally distributed by sex and grade.

Tabulation of the Results. —In the first survey a sex preference for titles was obtained simultaneously with the tabulation of the preferences for the total population. The tally marks were totaled for each title in each basal reader and a new list of titles, arranged in order of preference, was established for each reader. This data readily supplied a list of the ten most attractive titles in each basal reader and served as a basis for the second survey. In the second survey, a tabulation was obtained not only of the title and story preferences of the total population but for sex and intelligence differences as well. Only the L, Yes, and No choices were recorded as the results of the L, D, and Don't know columns were not needed for purposes of analysis.

#### Analysis of Data

Table I indicates the ten most attractive titles and the ten least attractive titles of the three intermediate grade basal readers published by Silver Burdett Company. It was compiled from data obtained in the first questionnaire. Similar lists have been established for the remaining basal readers and are available if desired. In this first study it is evident that percentages of children selecting stories is dependent upon the number of stories in a book. Books with fewer titles received higher percentages and as a result no cross comparisons can be made.

An analysis of the lists of least popular titles, indicates that the vocabulary is a factor in determining the popularity of a title. Many titles

which met with little enthusiasm contained words which were unfamiliar, meaningless, and foreign to the vocabulary of the children. Titles which are illustrative of these factors are: "Oxen for Anpu," "Kintu," "The Jar of Tassai," "Lotor the Washer," "Hoosier Barbecue," "Elijah the Fishbite," "Ti Leaf Slid in a Luau," "The Bombero of Guayaquil," "Demeter and Persephone," and "The Mazer of Lucrezia."

In contrast to the above, titles which employed familiar and meaningful vocabulary were selected much more frequently and were popular. Examples of these titles are: "Secret Cave," "Flying Snake and the Indians," "The Princess Who Could Not Cry," "The Goat that Went to School," "Dr. Doolittle and the Pirates," "Captured by the Indians," "Child of the Jungle," "Ali Baba and the Forty Thieves," "Pecos Bill and His Bouncing Bride," and "Gray Eagle's Horses."

The primary purpose of the second survey was to discover whether children select a title because of title appeal or story appeal. The data from this second questionnaire was based upon the 50 most highly rated titles from the five basal readers for each grade. Table II indicates the average of pupil preference for the ten best titles and stories for each of the intermediate grade basal readers published by the various companies.

It can readily be seen that the children emphatically preferred the stories to the titles in all of the basal readers involved in this survey. There were only five instances in the fifteen basal readers in which a higher percentage was noted for stories which had not been read than for stories which had been read. These titles were: "The Princess Who Could Not Cry," "The Sky Writer," "The Pony Express," "The Trial of the Rocket," and "Michael and Molly in the Movies." There were, however, some instances in which the percentage of preference for titles and stories showed less than a five percent difference, thus indicating that both story and title appealed to the children. Children in the fourth grade were enthusiastic about both story and title of: "The Princess Who Could Not Cry," "Dr. Doolittle and the Pirates," "A Famous Flight," "Little Trader of the Congo," and "The Floating House." Children in the fifth grade showed the same marked preference for both title and story of: "Red Blizzard," "The Pony Express," and "Captured by the Indians." The sixth graders enjoyed both the title and story of: "A Shooting Star that Hit the Earth," and "The Enchanted Island."

The second survey was also analyzed for sex differences and intelligence differences in title and story appeal. The data showed that although boys were unwilling to choose titles or stories which dealt with girls, the girls frequently selected titles and stories that appealed to the boys and in which the interest centered around adventure. "The Los Street Car," "Pony Express," "How Arthur Became King," "Down in Davy Jones's Locker," and "Indian Ponies," were chosen more frequently by the boys but were also liked by an almost equally high per-

TABLE I

PERCENTAGES OF CHILDREN SELECTING THE TEN MOST ATTRACTIVE TITLES AND THE TEN LEAST ATTRACTIVE TITLES OF THREE INTERMEDIATE GRADE BASAL READERS

Ten Most Popular Titles from Each of the Three Basal Readers of Silver, Burdett Company

DISTANT DOORWAYS Grade IV	FRONTIERS OLD AND NEW Grade V	ON THE LONG ROAD Grade VI
1. Flying Snake and the Indians 42.8	1. Captured by the Indians 60.6	1. A "Shooting Star" that Hit the Earth 44.5
2. The Maid with the Golden Hair 32.6	2. The Pony Express 41.1	2. The Romans 30.4
3. The Strawberry Goat 31.4	3. Arrow, the Alaskan Dog 35.9	3. Explorers of the Sky 28.2
4. Jungle Boy of the Amazon 26.8	4. The Trial of the Rocket 35.1	4. Stone Age Men in North American 26.8
5. On the Western Plains 26.6	5. Abe Lincoln's Cupboard 32.7	5. Michael and Mollie in the Movies 26.6
6. Kings of the Mountains 25.2	6. A Son of the Frontier 23.8	6. Two Young Mammoth Hunters 23.6
7. The Story Told by Bones 22.8	7. Mexico—Old and New 21.0	7. The Sky People 21.4
8. Little Trader of the Congo 17.4	8. Why the Wood Duck has Red Eyes 19.0	8. Tall Girl Saves the Charm 20.6
9. A Famous Explorer of Today 17.4	9. The Legend of the One-Eyed Ox 16.4	9. The House on Wheels 19.4
10. How Wild Animals Care for Their Young 16.4	10. The New and the Old in the Phillipines 15.2	10. Bob Meets the Stars 18.2

Ten Least Popular Titles from Each of the Three Basal Readers of Silver, Burdett Company

1. Oxen Anpu 1.0	1. Ti Leaf Sliding and a Luau 1.0	1. The Mazer of Lucrezia 1.2
2. A New Harness for Chiquito 1.2	2. Chestnut Burrs 2.0	2. Chants of the Field Workers 1.2
3. Seeds and Plants 1.8	3. Eduardo and John 2.2	3. Nakht at the Scribal School 2.0
4. Ned Digs in Iraq 2.0	4. Bird Wanderers 2.6	4. Caius and the Strange Boy from Britain 2.2
5. Finding a Way 2.2	5. Parade Ground 3.4	5. Pan Ku and the Beginning of Life 2.4
6. Demo, the Shepherd Dog 2.6	6. Mr. Fithian 3.6	6. The Guild Hall 2.4
7. With Vikings Bold 2.8	7. Some Tall Plants in the Phillipines 3.6	7. The Scroll-Maker 2.6
8. Animals Useful to Man 3.0	8. State House Square 4.2	8. Brother Soxmo's Garden 3.2
9. From the Bark of a Tree 3.2	9. The Woods Road 4.2	9. Nakht's Nile Journey 3.4
10. The Flower Hill Club 3.2	10. Our Bird Friends 5.4	10. Acinia's Home at Pompeii 4.0

TABLE II  
AVERAGE OF PUPILS LIKING OF TEN BEST TITLES AND STORIES IN  
BASAL READERS

Publisher	Grade IV			Grade V			Grade VI		
	Title %	Story %	Title %	Story %	Title %	Story %	Title %	Story %	
D. C. Heath and Company	47.8	66.8	58.9	76.4	45.0	79.9			
The Macmillan Company	49.4	57.9	51.0	77.1	49.3	85.8			
Scott, Foresman and Company	48.5	67.1	59.5	82.5	59.5	76.5			
Silver, Burdett Company	50.8	67.9	61.3	74.8	40.3	69.6			
Ginn and Company	46.2	63.3	53.7	76.5	46.6	82.2			

centage of girls. On the other hand, "The Princess Who Could Not Cry," "The Maid with the Golden Hair," "The Proud Princess," "The Girl Who Did What She Wanted," and "Tall Girl Saves the Charm," although very popular with the girls, were infrequently selected by the boys.

Boys expressed great interest in titles and stories which dealt with animals, the out-of-doors, adventure, exploration, and heroes. "Danger on the Docks," "How Wild Animals Care for Their Young," "Jungle Boy of the Amazon," "Captured by the Indians," "Wild Dog," "William Tell," "Gray Eagle's Horse," "Robin Hood," and "The Young Mammoth Hunters," are all stories which rated highly with the boys. Girls demonstrated a marked preference for stories which suggested the imaginative, the emotional, and homelife and familiar experiences. "The Strawberry Goat," "Seven Dancing Stars," "The Golden Eggs," "The Golden Touch," "The Silver Star," and "How Boots Befooled the King," are titles of stories which indicate the interests of the girls.

The second survey revealed that, in general, the gifted children selected stories that resembled the choices of dull children. Animal and fairy stories played an important role in the selections of both groups. "The Goat that Went to School," "The Legend of the One-Eyed Ox," "Wild Dog," "Wolf Pit," "Wild-Horse Roundup," and "Buffalo Hunt" are examples of popular choices in the animal category. Illustrations of imaginative stories which held appeal for both intelligence groups were: "The Steadfast Tin Soldier," "Saint George," "The Proud Princess," "Ali Baba and the Forty Thieves," "The King of the Golden River." On the other hand, differences between the upper and lower quartiles were apparent in stories which dealt with familiar experiences. Dull children expressed a greater degree of liking for such stories as "The Seventh Pup," "Bingo and the Angry Rooster," "Cast Away in a Department Store," "The Last Snake," "Michael and Molly at the Movies," and "Barnum's First Circus." Stories based on factual materials were favored by children in the upper quartile. Included in this area were "The Story Told by Bones," "Mexico—Old and New," "The Enchanted Island," "The Forty-Niners," "The Romans," and "The Stone-Age Men in North America."

### Conclusions

This investigation does not include all the basal readers of publishing companies that are available in the intermediate grade classrooms, but it has increased our knowledge of the preferences of children in grades IV, V, and VI for some basal reader titles and stories. It is apparent that the publishers have included many titles and stories in their basal readers that are very appealing and acceptable to the children. However, the marked unpopularity of some titles would indicate that publishers

might well submit titles to the children for their reaction and approval prior to their inclusion in the basal readers. It would seem advantageous, too, to eliminate from these readers, titles which include a feminine character because boys continually rejected titles suggesting a girl. Many of these stories avoided by the boys because of the title actually contained interest factors that would appeal to boys. Titles including meaningless, strange, and foreign words should be revised to encourage the reading of the story, because they, too, contained many elements of appeal to this age group.

# TRENDS AND ISSUES IN STANDARDIZED TESTING\*

HERBERT GLENN LUDLOW  
University of Michigan

TESTING TODAY is truly a "big business." A writer (1) for The American Psychologist in 1947 estimated that in one year 20,000,000 Americans took a total of 60,000,000 tests. Now, as with all big business, it becomes necessary to sit down periodically and examine the balance sheet. We may safely leave the actual financial statement to the comptrollers of the test publishers, but we measurement workers must assume responsibility for the construction and use of the technically complex tests themselves. In this article I should like to present a kind of airplane view of some of the important issues and trends in standardized testing. For the sake of convenience and brevity, I have grouped these thoughts under six headings; namely, the measurement lag, overspecialization, excessive individualization, intelligence testing, tests for drive, and technical considerations.

## The Measurement Lag

One of the most difficult issues facing the test constructor today is the creation of tests designed to measure the newer and more intangible objectives of modern education. As an example of this measurement problem, let us consider the field of arithmetic. Standardized tests for the measurement of computational ability and the solution of isolated verbal problems were developed quite early. However about ten years ago, the arithmetic curriculum people began talking and writing about the "meaning theory" of instruction. This approach emphasizes such goals as understanding, quantitative thinking, and number vocabulary in addition to the traditional aims. Obviously, this shift in instructional emphasis calls for action on the part of test experts in creating new types of items and tests to assess all the important outcomes of pupils' number experiences.

An analysis of the currently available commercial tests demonstrates that a real measurement lag exists in the field of arithmetic. Although there are a few indications of recognition and effort in the

\* Paper presented at a meeting of the National Council on Measurements Used in Education, St. Louis, Missouri, February 26, 1952.

assessment of these more recent mathematical aims, intensive work on the part of measurement technicians is called for. Arithmetic has long been considered to be a field well adapted to standardized testing. The urgency of the present problem stems partially from the viewpoints of some educators who feel that non-test techniques chiefly must be called into action in the case of the "new arithmetic." It may well be true that teacher observation, conference, and interview techniques will prove to be quite helpful. The point is, however, that the long and fruitful contribution of good arithmetic tests should not be disregarded. I believe the final outcome is up to the test technicians. What has just been said about arithmetic tests applies equally to many other fields of instruction where pupil understanding and insight is becoming the cardinal objective.

### Overspecialization

The measurement lag brings up another issue in regard to today's test experts. Some quite prominent educators are beginning to question the validity of the products of test constructors who evidence lack of information in the total instructional pattern. That is, critics are pointing out that the four basic steps in the instructional procedure, objectives, content, organization, and evaluation, are all quite intimately related and interdependent. Consequently, it seems that researchers primarily interested in test construction must cultivate a higher order of communication with curriculum and methods workers. Insofar as these other groups feel that we are unaware of the total educational problem, to that extent they appear to distrust our products. Thus it seems that measurement experts must attempt to understand even more thoroughly the thinking of educational philosophers and methodologists. We must never allow highly technical considerations of construction to blind us to the frontier thinking of our curriculum colleagues and the resulting shift of instructional emphases. Perhaps, as Robert L. Ebel (3) has recently said, in many cases the energy expended on correlational validation studies on the basis of unreliable criteria would have been spent better in a thorough analysis of the face value of the proposed instrument itself!

### Excessive Individualization

One of the most significant trends in the field of measurement research and test construction today is the movement toward cooperative endeavor. Our past history was characterized largely by individual effort, but the scope of modern evaluative problems seems to demand union and a more concerted approach in research and construction. For

an example of this trend, we may note the Educational Testing Service formed by the merging of the testing activities of the College Entrance Examination Board, the Cooperative Test Service, and the Graduate Record Examination. These three non-profit agencies operating at different levels had begun to overlap, thus duplicating research and operational costs. Although there are still plenty of intellectual frontiers to be explored by hardy individuals, the refinements and perfection of these endeavors may be more completely realized by joint action.

Further evidence of the trend away from excessive individualism may be noted in the publication of recent books in the field. The important American Council of Education publication, Educational Measurement, has some seventy collaborators. Buros' Mental Measurement Yearbooks afford another example of the remarkable results which may be attained by mutual activity.

### Intelligence Testing

Intelligence tests have enjoyed a high degree of acceptance and widespread usage. Few psychologists would deny that current group and individual intelligence tests are more valid and reliable than most other types of standardized instruments, such as aptitude tests and interest and personality inventories. Nevertheless, there is reason to believe that all is not well with most of our paper-and-pencil tests of mental ability. The Allison Davis Group at Chicago is becoming increasingly vocal in their assertion that the present tests are hopelessly culturally biased. These researchers are saying that many intelligence test validation studies have suffered from circular reasoning. That is, according to the usual procedure for checking internal consistency in the experimental forms of intelligence tests, the lowest third of pupils (chiefly from the lowest socio-economic groups) are not allowed to show in the final test, any of the activities at which they are equal or superior to the middle and higher class children. To remedy this situation, the Davis School proposes long range research to discover symbols and problems that are equally familiar and motivating to all social and economic groups. It is possible that the next decade will bring some significant changes in item selection techniques and construction of intelligence tests.

### Tests for Drive

Educators realize that a pupil's drives or motivating forces affect results on many kinds of tests. Yet, the very complexity of human behavior has retarded the efforts to assess these tremendously important sources of individual differences in test behavior. It seems

safe to predict that more research effort will be expended in this area and that some degree of success will be attained. Let's consider one example, again from the field of mathematics. E. I. Sawin's 1951 Ph.D. dissertation (4) presents a "Motivation Inventory for Mathematics." This novel instrument attempts to explore students' most intense interests and areas of most favorable or unfavorable motivations in regard to mathematics. The inventory is a long one consisting of 319 items, divided into three types. The following is an example:

Imagine yourself in the following situation: One of your classmates is telling you that he (or she) has been working on an idea for a way to extract square root which is better than the one commonly used. He (or she) wants to explain it to you, to have you check it over to see if the reasoning is correct, and to try it out to see how it works.

25. Would you be interested in hearing the idea explained?

- A. Yes
- B. Uncertain, but probably Yes
- C. Uncertain, but probably No
- D. No.

One interesting finding was that the correlation coefficients between scores on the motivation inventory and various other measures of pupil ability and adjustment were all positive. Sawin concluded that this means that an increase in favorable motivation for mathematics is associated with an increase in ability or adjustment. Although the reader perceives certain weaknesses in this approach to the assessment of drive and interest, he will probably admit that Sawin's study contains implications for future research in the measurement of motivation.

#### Technical Considerations

Although it borders on the ridiculous to treat the many recent technical considerations of testing in a brief section, the writer who fails to even mention this point would be guilty of gross neglect. World War II drove home the necessity for devoting much time and thought to the bases of validation. Psychological lessons from the last great conflict emphasize the importance of describing our criterion completely and investigating its adequacy as an indication of success. In fact, test experts are now talking about three categories of criteria, immediate, intermediate, and ultimate. Further, the lag between the time when tests are given and the time when criteria of success are available is receiving more attention. Longitudinal validation studies

should and probably will become more numerous. Finally, the necessity for cross-validation is being realized. Although this principle is an old one, it is only fairly recently that test producers have begun to expend the time, money, and energy necessary for effective cross-validation studies.

The factor analysis approach to test construction received a real impetus through the Air Force psychological research program of World War II. The proponents of the various brands of factor analysis believe that this technique makes possible the refinement and purification of tests, and reveals what kinds of tests may be developed. In spite of the limitations and criticisms of factor analysis, it appears that this technique will continue to be exploited. Undoubtedly, debate concerning its values and functions will persist.

Reliability has not been neglected in the recent literature. While too many test consumers still speak of "the" reliability coefficient, we recognize now three types of reliability coefficients: the coefficient of stability, the coefficient of equivalence, and the coefficient of stability and equivalence. We know that the split-half and Kuder-Richardson methods must not be used for speeded tests. Unfortunately this principle is sometimes disregarded and the unsophisticate is misled by a grossly inflated coefficient of reliability. In view of the relative complexity of the reliability concept, test users should be reeducated in regard to the importance of the stability of test scores. Personnel selecting tests must be made aware of the fact that there is no one standard for an adequate reliability coefficient.

Cronbach (2) has shown recently that the influence of response sets on test scores is not adequately appreciated. He has stated that the only major form free from response set seems to be the multiple-choice type. Inasmuch as response set tends to reduce the validity of test scores, one can only wonder that this phenomenon has escaped consideration for so long.

In summary, recent research on validity, reliability, factor analysis, and response set has produced results which will surely improve the science of test construction. However, one cannot resist speculating whether these statistical considerations have outrun the quality of the tests themselves.

### Conclusion

The next decade promises to be a critically important one for standardized testing. Although we have long since stopped regarding standardized testing as a panacea for all educational ills, it seems safe to forecast that this form of assessment will continue to play a strong role in all worthwhile evaluative programs.

If those who object to quantitative thinking in education will set themselves to work to understand it; if those who criticize its presuppositions and methods will do actual experimental work to improve its general logic and detailed procedure; if those who are now at work in devising and in using means of measurement will continue their work, the next decade will bring sure gains in both theory and practice. Of the gains made in the last decade, we may well be proud. (5)

This well-known quotation of Edward L. Thorndike in 1918 seems to summarize quite adequately our present feeling of pride in past accomplishments and our attitude of confidence in the future.

#### REFERENCES

1. Anonymous, "Testing is Big Business," The American Psychologist, (1947), 2, 26.
2. Cronbach, Lee J., "Further Evidence on Response Sets and Test Design," Educational and Psychological Measurement, Vol. 10, No. 1 (Spring, 1950), 3-31.
3. Ebel, Robert L., "Construction and Validation of Tests," Chapter IX, in Review of Educational Research, Vol. XX, No. 1 (February, 1950), p. 93.
4. Sawin, E. I., "Motivation in Mathematics: Its Theoretical Basis, Measurement, and Relationships with Other Factors," The Mathematics Teacher, Vol. XLIV, No. 7, (November, 1951), p. 471-478 (Abstract of Ph. D. Dissertation, U. of Chicago, 1951).
5. Thorndike, Edward L., "The Nature, Purposes and General Methods of Measurements of Educational Products," Seventeenth Yearbook of the National Society for the Study of Education, Part II. Bloomington, Illinois: Public School Publishing Company (1918). Chapter II, p. 24.

# A LONGITUDINAL COMPARISON OF FRESHMAN-SENIOR STANDING

LESLIE C. CARLIN

Central Michigan College of Education  
Mount Pleasant, Michigan

## Introduction

TODAY THE giving of tests and inventories for guidance purposes is common practice among colleges and universities. The choice of the instruments to be used presents a difficult problem. This report\* is a contribution to its solution.

This study differs somewhat from many similar studies. The data used were, (1) percentile ranks achieved when the students were freshmen, and (2) the cumulative scholastic average that each student had when he graduated. The population consisted of 312 students who entered college in 1946 and who graduated in 1950.

## Procedure

A five by five analysis was used to determine the relationship between the percentile ranks and cumulative scholastic averages in each subject. The five by five analysis consisted in dividing the grades earned by students into five categories, as follows: A - superior grade, five grade points per credit hour; B - above average, four grade points per credit hour; C - average, three grade points per credit hour; D - below average, two grade points per credit hour; and E - failure, one grade point per credit hour. The averages were then divided according to the normal distribution. This yielded the following: 1.0-1.7, group one; 1.8-2.6, group two; 2.7-3.4, group three; 3.5-4.2, group four; and 4.3-5.0, group five. On the same basis the percentile ranks were divided as follows: 1-7, group one; 8-30, group two; 31-69, group three; 70-92, group four; and 93 and above, group five.

The twenty-five cells, created by the application of the five by five analysis, revealed the number of students who ranked in each percentile group and whose scholastic cumulative average fell in each grade grouping. The coefficient of contingency was computed for each set of distributions. The formula for this coefficient may be stated as:

\*Partial findings of the author's Field Study No. II for Doctor of Education Degree, Colorado State College of Education, Greeley, June 1951.

$$C = \sqrt{\frac{T - 1}{T}}$$

In this, T represents  $\Sigma \frac{(f^2_{rc})}{(f_r f_c)}$ , in which  $f_{rc}$  is the frequency in a given cell of the table,  $f_r$  the total frequency in the row in which the cell lies, and  $f_c$  that in the column.

The five by five analysis was used to determine the relationship between the average grades received in English, foreign language, agriculture, biology, chemistry, geography, mathematics, psychology (educational and general), economics, history, political science, sociology, commerce, home economics, industrial arts, physical education, art, music, directed teaching, education, and total college scholarship; and the percentile scores on the American Council on Education Psychological Examination, College Edition; American Council on Education Cooperative English Test, Higher Level; and California Test of Personality, Grades 9-College. Also, the following were used when they were applicable: American Council on Education Cooperative General Achievement Test in Mathematics; Natural Science; and Social Science.

#### Description of Tests

The American Council on Education Psychological Examination yields two part scores: one designated the Q score, and the other the L score. According to the authors the former indicates aptitude in arithmetical reasoning. The latter reveals linguistic aptitude. One of the main purposes of this examination is to appraise what has been called scholastic aptitude, with special reference to the requirements to most high school and college curricula.

The achievement tests furnish measures of the educational accomplishments of the individual. They show levels of attainment in various fields. They sample knowledge of facts and principles, and of certain skills.

The ACE Cooperative English Test measures Reading Comprehension, Mechanics of Expression, and Effectiveness of Expression.

Knowledge of the sciences is furnished by the ACE Cooperative General Achievement Test in Mathematics; ACE Cooperative General Achievement Test in Natural Science; and ACE Cooperative General Achievement Test in Social Science.

The "California Test of Personality" is not a test at all despite its name, but it is a rather cleverly constructed questionnaire which has fifteen items constituting each of 12 sections. The first six sections, which includes self reliance, personal worth, personal freedom, feeling of belongingness, freedom from withdrawing tendencies, and freedom

TABLE I

## SUMMARY OF DISTRIBUTION OF GRADE POINT AVERAGES AMONG SUBJECTS STUDIED

Subject	No. Taking Courses	Percent of Class	1.0	1.8	2.7	3.5	4.
			-	-	-	-	-
<b>Language and Literature</b>							
1. English	312	100.0	.6	4.5	49.4	40.4	5.
2. Foreign Language	104	33.0	2.9	12.5	37.5	29.8	17.
<b>Science</b>							
3. Agriculture	28	9.0	0.0	3.6	53.5	39.3	3.
4. Biology	157	51.0	1.9	9.6	38.2	39.5	10.
5. Chemistry	89	29.0	5.5	21.3	39.4	25.8	7.
6. Geography	141	45.0	1.4	10.6	43.3	38.3	6.
7. Mathematics	192	62.0	3.6	19.8	45.8	26.1	4.
8. Psychology	289	93.0	.3	5.2	36.7	43.6	14.
<b>Social Science</b>							
9. Economics	133	43.0	0.0	9.8	51.9	31.6	6.
10. History	249	80.0	0.0	10.8	42.2	43.0	4.
11. Political Science	308	99.0	0.0	13.0	55.8	26.3	4.
12. Sociology	221	71.0	0.0	5.4	56.1	29.4	9.
<b>Practical Arts</b>							
13. Commerce	98	31.0	0.0	4.1	42.8	48.0	5.
14. Home Economics	33	11.0	0.0	0.0	30.3	60.6	9.
15. Industrial Arts	99	32.0	0.0	9.1	56.6	31.3	3.
<b>Physical Education</b>							
16. Physical Education	92	29.0	0.0	1.1	30.4	63.1	5.
<b>Fine Arts</b>							
17. Art	89	29.0	0.0	3.4	49.4	42.7	4.
18. Music	77	25.0	0.0	3.9	36.4	41.5	18.
<b>Education</b>							
19. Directed Teaching	237	76.0	0.0	0.0	4.2	55.7	40.
20. Education	241	77.0	0.0	2.9	28.2	51.5	17.
<b>Total</b>							
21. Total College Scholarship	312	100.0	0.0	0.0	44.9	50.3	4.

TABLE II

RELATIONSHIP BETWEEN PERCENTILE RANKS ACHIEVED BY FRESHMEN AND CORRESPONDING SCHOLASTIC AVERAGE ACHIEVEMENT BY END OF SENIOR YEAR

Subject	No. of Cases	L Score	Q Score	English	Self Adjustment	Social Adjustment	Achievement Tests
<b>Language and Literature</b>							
1. English	312	.42	.24	.54	.32	.42	
2. Foreign Language	104	.49	.49	.39	.40	.46	
<b>Science</b>							
3. Agriculture	28	.63	.46	.61	.35	.42	
4. Biology	157	.43	.27	.42	.40	.46	.41
5. Chemistry	89	.40	.42	.36	.43	.42	
6. Geography	141	.39	.32	.32	.34	.34	
7. Mathematics	192	.30	.41	.35	.32	.29	.44
8. Psychology	289	.39	.21	.38	.32	.27	
<b>Social Science</b>							
9. Economics	133	.48	.44	.50	.26	.27	.49
10. History	249	.38	.29	.41	.20	.20	.35
11. Political Science	308	.36	.29	.29	.20	.12	.40
12. Sociology	221	.35	.27	.34	.20	.33	.55
<b>Practical Arts</b>							
13. Commerce	98	.30	.35	.42	.26	.25	
14. Home Economics	33	.36	.54	.42	.49	.51	
15. Industrial Arts	99	.33	.40	.35	.49	.43	
<b>Physical Education</b>							
16. Physical Education	92	.38	.36	.10	.35	.30	
<b>Fine Arts</b>							
17. Art	89	.44	.41	.38	.18	.32	
18. Music	77	.44	.40	.44	.31	.21	
<b>Education</b>							
19. Directed Teaching	237	.24	.24	.21	.41	.19	
20. Education	241	.43	.32	.38	.14	.20	
<b>Total</b>							
21. Total College Scholarship	312	.42	.29	.45	.20	.30	

from nervous symptoms, yields a total self adjustment score while the last six sections which includes social standards, social skills, freedom from anti-social tendencies, family relations, social relations, and community relations, gives a total social adjustment score. It was the two total scores which were used in this study.

### Results

In Table I is presented a summary of grade point average distribution for the courses taken by the students. The table reveals the total number of students who took each course; the percent of the class each total represents; and the grade point average distribution, in percent, for each course. Also included is the grade point average distribution for total college scholarship.

Table II gives a detailed comparative picture of the relationships existing between the percentile ranks established by the students when freshmen on tests and questionnaires and their average grade points earned in each course being considered. Also included is their average total college scholarship.

### Summary and Conclusions

1. The linguistic score on the ACE Psychological Examination is a definitely superior basis for predicting achievement in college English and literature courses than the quantitative score on the same test, but the ACE Cooperative English test score indicates even better the probabilities of achievement in these areas than does the linguistic score.
2. It is interesting to note that no appreciable difference exists between the linguistic score and the quantitative score on the ACE Psychological Examination in their relationship with grades made in college foreign language courses.
3. The idea that the quantitative score on the ACE Psychological Examination indicates a greater degree of success in scientific courses than does the linguistic score on the same test is not substantiated. The quantitative score is the better indicator of achievement in college mathematics courses. The linguistic score is just as good as or a better predictor of success in agriculture, biology, chemistry, geography, and psychology.

4. Seemingly, there is no appreciable advantage in retaining the use of the ACE Cooperative General Achievement Tests in mathematics or natural science.

5. The best overall predictor of achievement in the social sciences is the ACE Cooperative General Achievement Test in Social Science score.

6. The correlations established between the test scores and achievement in the practical arts, which includes commerce, home economics, and industrial arts, do not establish a pattern which accentuates any one test as being significantly superior from the predictive viewpoint.
7. The relationship between the ACE Psychological Examination linguistic score and quantitative score and teachers' grades in art and music indicates that it takes as much academic aptitude to do well in the fine arts as it does in the academic fields.

# CLASSROOM CHOICE STATUS RELATED TO SCORES ON COMPONENTS OF THE CALIFORNIA TEST OF PERSONALITY

ONAS C. SCANDRETTE

South Kitsap Junior-Senior High School  
Port Orchard, Washington

SINCE THE PIONEERING WORK of Moreno (7) and Jennings (4) many sociometric studies have been made. Most of these studies have used elementary, secondary, and college students as subjects. Representative studies have shown a positive relationship to exist between choice status and the following factors: propinquity (8, 11); frequency of association (3, 9, 10); age (3, 12); size (3, 12); intelligence (3, 12); maturity (3); scholarship (12); parental occupation (8); similarity of interests (2); and personality traits (1, 2, 5).

Personality data have been obtained by asking students to rate one another on such traits as "cheerful," "friendly," "enthusiastic," etc.; by having students list the qualities they admire in persons chosen as best friends; and by the use of personality tests. The last method has been used less frequently than the other two. Two previous studies (1, 6) have made use of the California Test of Personality. Neither of these studies is strictly comparable to the present study, however, since different choice criteria were used. The criterion of choice in Sister Alexandra's study (1) was election to scout, church or school offices. Choice status in the present study was determined by frequency of choice as a fellow worker on a classroom committee. It is quite possible that persons chosen as fellow workers would not have the same personality characteristics as those chosen as leaders. This fact should be kept in mind in making comparisons between the two studies. In the study by McClelland and Ratcliff (6) choice status was determined by asking students whom they would choose as companions for four varied types of activities.

## Methodology

Subjects of the present study were 68 eighth grade students from two social science classes at South Kitsap Junior High School, Port Or-

TABLE I

MEAN LIFE ADJUSTMENT SCORES OF FREQUENTLY AND  
INFREQUENTLY CHOSEN STUDENTS

	Five or More Choices (N=24)	One or no Choice (N=18)
Mean	141. 41	118. 78
Range	113-171	72-156
Standard Deviation	15. 12	25. 64
Standard Error of Mean	3. 15	6. 22
Critical Ratio	3. 25	

chard, Washington. The study was made during the 1950-51 term. Choice data were obtained by asking each student to list the names of two boys and two girls with whom he would like to work on a social science project committee. The number of choices received by individual students ranged from 0 to 13. It was originally intended to compare the personality adjustment scores of those receiving no choices with an equal number of students receiving the highest number of choices. As only seven students received no choices, however, it was decided to compare those receiving one or no choice with those receiving five or more choices. This division placed 18 in the infrequently chosen group and 24 in the frequently chosen group.

At a later date, all students were given the Intermediate Series, Form A, of the California Test of Personality. This test is divided into 12 separate components, each component consisting of 15 "yes-no" questions. Components purportedly test self reliance, sense of personal worth, sense of personal freedom, feeling of belonging, freedom from withdrawing tendencies, freedom from nervous symptoms, social standards, social skills, freedom from anti-social tendencies, family relations, school relations, and community relations. The California Test of Personality yields a total life adjustment score (180 points possible) and a score for each of the components (15 points possible).

### Findings

Table I shows that the mean life adjustment score of students receiving five or more choices was 141.41 as compared with 118.78 for those receiving one or no choice. A critical ratio of 3.25 indicates that this difference is statistically significant.

In her study of "Personality Adjustment and Leadership," Sister Alexandra (1) found a somewhat smaller but significant difference in favor of the leader group. McClelland and Rataliff (6) found that pupils who received the most acceptance choices had higher average personality test scores than those who received no choices.

Table II shows the mean score differences of the California Test of Personality. It will be noted that mean scores for all components except social standards, where the difference is negligible, are in favor of the more frequently chosen students. For only five of the components, however, is the difference of sufficient magnitude to be significant. It will be noted that four of the six self-adjustment component score differences were statistically significant as compared with only one of the six social adjustment components. Significant self adjustment differences exist for sense of personal worth, feeling of belonging, freedom from withdrawing tendencies, and sense of personal freedom.

TABLE II

MEAN SCORES OF FREQUENTLY AND INFREQUENTLY CHOSEN STUDENTS ON  
 ↗ COMPONENTS OF THE CALIFORNIA PERSONALITY TEST

	Mean			SD			S.E. of M.		
	FC*	IC**	FC	IC	FC	IC	C. R.		
<b>Self Adjustment</b>									
Self Reliance	8.63	8.28	2.55	3.02	.47	.73	.40		
Sense of Personal Worth	11.04	8.67	1.92	2.79	.40	.67	3.04		
Sense of Personal Freedom	13.67	10.83	1.70	3.84	.35	.93	3.73		
Feeling of Belonging	13.04	10.28	1.54	3.16	.32	.77	3.33		
Freedom from Withdrawing Tendencies	11.67	8.17	2.33	3.38	.49	.82	3.76		
Freedom from Nervous Symptoms	10.21	9.22	2.93	3.88	.61	.94	.88		
<b>Social Adjustment</b>									
Social Standards	13.04	13.11	1.19	1.76	.25	.42	.14		
Social Skills	11.00	10.00	2.40	2.58	.50	.63	1.23		
Freedom from Anti-Social Tendencies	11.33	9.39	2.11	3.34	.44	.81	2.11		
Family Relations	13.00	11.78	1.89	3.36	.39	.81	1.36		
School Relations	12.46	9.22	1.74	3.17	.37	.77	3.81		
Community Relations	12.21	11.61	2.81	2.24	.59	.54	.75		

\* Frequently Chosen (five or more choices)

\*\* Infrequently Chosen (one or no choice)

Leaders in Sister Alexandra's study had higher mean scores on each of the self adjustment components than non-leaders but the differences were significant for only three components—self reliance, sense of personal worth, and freedom from nervous symptoms.

In the present study the only social adjustment component which significantly differentiated between the most and least frequently chosen students was that of school relations. Although not statistically significant, all differences in the other social adjustment components, with the exception of "social standards," were in favor of the most frequently chosen group.

The results of Sister Alexandra's leadership study were quite different. She found mean differences in favor of the leaders for social skills, freedom from anti-social tendencies, and family relations. Mean differences in favor of the non-leaders were found for social standards, school relations and community relations. The only two statistically significant component score differences, however, were social standards and school relations, both in favor of the non-leaders.

### Summary

1. Mean total scores on the California Personality test differentiated clearly between the most and least frequently chosen students, the mean score of the former being 141.41 as compared with 118.78 for the latter. The critical ratio was 3.25.

2. All but one of 12 components of the test revealed differences in favor of the most frequently chosen group. The one exception was social standards in which difference, though negligible, was in favor of the least frequently chosen group.

3. The mean differences between the scores of the most and the least frequently chosen students for four of the six self adjustment components were statistically significant. The four components which significantly differentiated the two groups were sense of personal worth, sense of personal freedom, feeling of belonging, and freedom from withdrawing tendencies.

4. Only one of the six components of social adjustment, school relations, significantly differentiated between the most and least frequently chosen students. Since choices were made for fellow committee members in a classroom situation it seems logical that those who were best adjusted to school life should be most frequently chosen.

5. With the exception of one component, school relations, scores made on self adjustment components differentiated more clearly between most and least chosen students than did scores of the social adjustment components.

## REFERENCES

1. Alexandra, Sister Mary, "Personality Adjustment and Leadership," Education (1946), 66, 584-590.
2. Austin, M. C. and Thompson, G. G., "Children's Friendships: A Study of Bases on Which Children Select and Reject Their Best Friends," Journal of Educational Psychology (1948), 39, 101-116.
3. Fursey, P., "Some Factors Influencing the Selection of Boy's Chums," Journal of Applied Psychology (1929), 11, 47-51.
4. Jennings, Helen Hall, Leadership and Isolation, A Study of Personality in Inter-Personal Relations, New York: Longmans, Green and Co.
5. Kuhlen, R. G. and Lee, B. J., "Personality Characteristics and Social Acceptability in Adolescence," Journal of Educational Psychology (1943), 34, 321-340.
6. McClelland, F. M. and Ratcliff, John A., "The Use of Sociometry as an Aid in Promoting Social Adjustment in a Ninth Grade Homeroom," Sociometry (1947), Volume 10, number 2.
7. Moreno, J. L., Who Shall Survive? Washington, D. C.: Nervous and Mental Diseases Publishing Company (1943).
8. Potashin, R., "A Sociometric Study of Children's Friendships," Sociometry (1946), 9, 48-70.
9. Reader, N. and English, H. B., "Personality Factors in Adolescent Friendships," Journal of Consulting Psychology (1947), 11, 212-220.
10. Scandrette, Onas C., "Friendship in Junior High Seventh Grades, A Study in Adjustment," Clearing House (1951), 25, 364-366.
11. Seagoe, M., "Factors Influencing the Choice of Associates," Journal of Educational Research (1926), 14, 126-132.
12. Wellman, B., "The School Child's Choice of Companions," Journal of Educational Research (1926), 14, 126-132.

# PRIMARY MENTAL ABILITIES AND OCCUPATIONAL CHOICES

LOUIS G. SCHMIDT  
Indiana University

TO WHAT extent are one's test performances related to the type of occupation he chooses? The answer to this question is important from the standpoint of its bearing on the theory of mental measurement and its importance for vocational and educational guidance. Guidance workers long have looked for instruments that would point in a short time specific occupations for individuals to follow. Probably nothing prevents a healthy outlook more in the field of guidance than to say that a pupil is cut out for some specific career and that in no other will he find equal success and satisfaction. Yet this idea, as fostered by many commercial testing agencies, may have unpleasant consequences for the individual who has set for himself a too restricted goal. Stimulation and satisfaction can usually be found in more than one specified job.

A person who is not a good student but who wishes to become a doctor, may find happiness as a hospital attendant, receptionist, or as a worker for a drug firm. It would make counseling more effective if adequate interest inventories were available as a supplement to performance tests but at their present stage of development, they cannot be made, as they often are, the focal point of vocational counseling. Measures of ability are safer, but relatively very little research has been completed in this area. An investigation of the value of ability test performance in counseling is reported below.

In this investigation the test of Primary Mental Abilities published by Science Research Associates, was selected because of the reputations of its authors and the fact that it gave five factor scores that were said to be directly related to many selected occupations. The manual furnished with this test indicates that persons making high scores in any one factor probably would be best fitted for a particular occupation said to need that ability. A high score on the space factor, for example, purports to indicate the desirability of an occupational choice of machinist, pilot, or engineer.

Space (S) is the ability to think about objects in two or three dimensions. Blueprint reading, for example, requires this ability. The designer, electrician, machinist, pilot, engineer, and carpenter are typical workers who need ability to visualize objects in space. S is helpful in geometry, mechanical drawing,

art, manual training, physics, and geography classes. \*

If the above statement is true, it would indeed be a valuable aid to counselors. In this study we set out to find whether students who select various occupations actually scored high in the factors said to be necessary to perform that occupation.

Nine hundred students, the entire tenth grade classes (456 girls and 444 boys) of each of four Wisconsin high schools, were used as subjects in this study. The average chronological age of the experimental group was fifteen years and five months on September 1, 1948. The four schools which they attended were those used in the Wisconsin Counseling study designed to study the predictability of post-high school academic and occupational performances of representative high school youth. The study was sponsored by the Research Committee of the University of Wisconsin.

Occupational choices of the subjects were obtained at three different times during the first year in which the experiment was conducted. At the time each subject took the Henmon-Nelson Test in October of 1948, again during a thirty minute interview with each subject, and finally after each subject took the Primary Mental Abilities Test in April of 1949, the subjects occupational choices were obtained. All three occupational choices for each subject were compared to determine stability of preferences. They were considered stable if all three were identical or if the one stated during the interview matched one of the other two. All other combinations of choices were labeled "no choice" for purposes of this experiment. Eighty-three percent of the experimental group were found to have stable occupational choices according to these criteria.

The ten occupations selected most frequently were secretarial, teaching, nursing, artist, machinist, auto mechanic, doctor, draftsman, engineer, and bookkeeper. Although farming was the third most popular selection, it was dropped from consideration in this study because of the wide variety of jobs and skills represented in that occupation.

To obtain a consensus of estimates about the performances said to be desirable in each of the ten selected occupations, four sources were used. The first of these was a study of the ten occupations listed under each ability in the Primary Mental Abilities manual; second was an analysis of performances listed in the United States Employment Service General Aptitude Test Battery; third was an examination of job analyses made by the United States War Manpower Commission for each of the ten occupations, and fourth by scrutiny of the "Abilities Required" sec-

\*L. L. Thurstone. Manual of Instructions for the Primary Mental Abilities Test (Chicago: Science Research Associates, 1947).

tion of the Occupational Abstracts, published by New York University.

Results of the study were obtained from the application of three basic techniques. The first procedure was that of comparing primary mental abilities subtest scores of subjects who chose one of the ten most common occupations, with the scores said to be important for their occupational choices. The data were next analyzed to determine the relationship of the Henmon-Nelson Test of Mental Ability scores to the occupational choices of subjects who had achieved significantly high or low scores on individual subtest on the Primary Mental Abilities Test. Finally, a study was made of the tenability of the hypothesis that subjects tended to choose occupations which were said to require the test proficiencies they had achieved.

Study of the significant differences between scores made by subjects who selected certain occupations and the scores of those who did not reveals that:

1. Subjects selecting secretarial work as an occupation showed superiority in the verbal and reasoning subtests, but not on the word-fluency subtest. Examination of the literature in the sources given above suggested that verbal reasoning and word fluency were desirable qualifications for this occupation.
2. Subjects selecting teaching as an occupation made scores on the verbal and reasoning subtest. As above, verbal reasoning and word fluency had been said to be desirable for this occupation.
3. Subjects selecting nursing as an occupation made significantly higher than average scores on only the reasoning factor. Word fluency had also been indicated as desirable on the criteria given above.
4. Subjects selecting art as an occupation did not make significantly higher than average scores on the space subtest.
5. Spatial and number ability are described in the test manual as being important for machinists, but our subjects who selected machinists as an occupation failed to make scores significantly different from the average of all other groups on either subtest.
6. The same conditions as in number five above applied to subjects selecting auto mechanic as an occupation.
7. Subjects selecting doctor as an occupation failed to make significantly high scores on the space and reasoning subtest that were said to measure abilities desirable for doctors.

8. Subjects selecting drafting as an occupation showed superiority on the space subtest, but not on the number and reasoning subtests. All three were said to be desirable abilities for draftsmen.
9. Spatial and number abilities were said to be desirable for engineers, but subjects selecting engineering as an occupation failed to make significantly higher scores on those subtests.
10. Subjects selecting bookkeeping as an occupation made significantly high scores on the number subtest but not on the reasoning subtest. Both were said to be desirable abilities for bookkeepers.

In five of the ten occupations chosen, the subjects who selected them made significantly higher than average scores on subtest purporting to measure abilities that were said to be important in these occupations.

The coefficient of correlation between the total raw scores on the Primary Mental Abilities Test and the Henmon-Nelson Test of Mental Ability with a population of 900 was found to be .68. There is, therefore, some evidence that scores on the Primary Mental Abilities test provide information similar to that provided by scores on the Henmon-Nelson Test of Mental Ability. The coefficient is not high enough, however, to make accurate prediction of an individual's performance on one of these tests from the knowledge of the other.

It is concluded that: (1) Occupational choices of tenth grade secondary students are more stable than previous literature in this area indicates; and (2) sophomores in high schools who select certain specified occupations tend to make significantly higher than average scores on only approximately half of the tests of Primary Mental Abilities which purport to measure abilities required in those occupations.

# DO WE NEED A NEW BASIC VOCABULARY FOR HIGH SCHOOL LATIN

CAROLYN E. BOCK

University of Minnesota (Duluth Branch)  
Duluth, Minnesota

DO WE NEED a new or altered or implemented basic vocabulary in high school Latin? What purposes should a basic vocabulary serve? Are these purposes consistent with the overall objectives of the study of Latin? What is the existing basic vocabulary for high school Latin? What purpose is the College Board List designed to serve? One question at a time, and we'll save the first question for the last.

In the Classical Investigation of 1924, 98 percent of the teachers of secondary school Latin judged the ultimate objective of increased ability to understand the exact meaning of English words derived directly or indirectly from Latin, and increased accuracy in their use as valid for the course as a whole. Growth and development of English vocabulary is then a worthy objective of the study of Latin as expressed in the opinion of the teachers. It follows that this can be best achieved through the Latin vocabulary at the high school level. The degree of success in amount of transfer and gains made resulting from a study of Latin has been measured in controlled experimentation conducted by Greenough and Kittredge, Douglass and Kittelson, Thorndike and Ruger, Pond, Otis, Harris, Hamblen, and Haskell, to mention only a few. Reports of their investigations indicate that the range of comprehension by high school students of English words of Latin derivation was all the way from negligible to considerable. The extent and significance of the Latin element in our language are generally known. Is it not natural, therefore, to expect that students of Latin, who are doubly exposed to that element in our language, should far surpass those students who must acquire their knowledge of that element incidentally? The crux of the problem appears to be whether or not the vocabulary of high school Latin contributes to a representative English vocabulary, and whether it is adequate as a means for the development of a better understanding of English words of Latin derivation.

The Latin Word List, more commonly referred to as the College Board List, prepared by the College Entrance Examination Board, is the accepted basis of high school Latin vocabulary. The choice of words included in the College Board List is based upon the examination of so large an amount of Latin literature that the use of the list is not likely to restrict greatly the reading in the schools. The method of choice gave a preference to those words that are most useful, both in reading

at the specified stage and in that to be done later, with a slight preponderance of those words which occur most frequently in the works that have heretofore formed the reading in our schools. The list will serve to reassure teachers that deviation from the beaten path is safe provided they take the vocabulary as one of their guides. As a result of this last statement, it has been a sine qua non with all the writers of Latin textbooks in preparing vocabulary and reading selections for the high school Latin student. The authors are not restricted to the College Board List; but it is fundamental and constitutes the core vocabulary, with not too much variation in the vocabulary, of the textbooks in existence today.

In order to investigate what relationship exists between the high school Latin vocabulary and some standard measure of English vocabulary, a study was made, the details of which will be related briefly. As a measure of English vocabulary, the first and tenth words, according to eligibility principles previously established, on each page of the Thorndike Senior Century Dictionary made up the sample—a total of 4118 words. Of this number, 1916 words were derived from Classical Latin. The source of these words was checked against the College Board List, and the year to which the word was assigned in the College Board List from which the word was derived was noted.

The College Board List was examined to determine how many of its words were used in building English words appearing in the sample.

The findings are as follows:

A. Related to the Sample of English Words

1. Of the 1916 words in the sample derived from Classical Latin, 988 words or 51.56 percent were derived from Latin words assigned to the vocabulary of first and second year Latin in the College Board List.
2. Of the 1916 words in the sample derived from Classical Latin, 438 words or 22.91 percent were derived from Latin words assigned to the vocabulary of third and fourth year Latin in the College Board List.
3. Of the 1916 words in the sample derived from Classical Latin, 1427 words or 74.47 percent were derived from Latin words included in the vocabulary of the College Board List.
4. Of the 1916 words in the sample derived from Classical Latin, 489 words or 25.52 percent were derived from Latin words not included in the vocabulary of the College Board List.
5. Of the 1916 words in the sample derived from Classical Latin, 274 words were Latin compounds made up of elements assigned to

the vocabulary of different years of high school Latin by the College Board List and not in the list.

- a. One hundred sixty-five of the English words in the sample were derived from Latin compounds, one element of which was assigned to the vocabulary of first and second year Latin in the College Board List and another element from a word appearing in the vocabulary of third and fourth year Latin in the College Board List.
- b. Ninety-eight words in the sample were derived from Latin compounds, one element of which was from a Latin word assigned to the vocabulary of first and second year Latin and another element from a word not included in the College Board List.
- c. Three words in the sample were derived from Latin compounds, a part of which belonged to the vocabulary of third year Latin and another part of which belonged to the vocabulary of fourth year Latin in the College Board List.
- d. Six words in the sample were derived from Latin compounds, a part of which was from a Latin word allotted to the vocabulary of third and fourth year Latin in the College Board List and another part of which was from a word not occurring in the College Board List.

#### B. Related to the College Board List

1. Of the 1880 words making up the College Board List, 1135 words in the list did not yield any English derivatives included in the sample of 1916 words derived from Classical Latin. It should be noted, however, that the sample is a small fraction of the Thorndike list.
2. The 1916 English words in the sample were derived from 745 Latin words in the College Board List. Latin prepositions which became English prefixes were used most often in the formation of the 1916 words.
3. Three hundred eighty-one words were used only once in forming English words in the sample while the first five words in frequency were used 390 times to form that many English words.

### Conclusions

Based on the findings heretofore stated, one may arrive at certain conclusions concerning the College Board List and its contribution to English vocabulary. It is well to bear in mind constantly that these conclusions are drawn with the realization that the College Board List was designed to provide a vocabulary for the authors to read rather than for the purpose of teaching English words of Latin derivation. These conclusions are valid insofar as they do not interfere with the attainment of the immediate objective of Latin study, which is the ability to read and understand the selections presented for study.

1. The College Board List does not consider Latin's contribution to English word building and borrowing as its primary purpose; hence many of the values are neglected which might be realized if the List were set up with that aim.
2. Inasmuch as about 50 percent of the Latin derived English words contained in the sample come from Latin words in the first two years of the College Board List, it is apparent that the majority of high school Latin students who discontinue the study at the end of two years would have no experience through their Latin with words assigned to the third and fourth years.
3. If the mastery of Latin words in the College Board List yielding English words in the sample were absolute and perfect on the student's part, he would have access to half of the Latin derived sample at the end of two years study, and three fourths of the sample upon completion of four years of high school Latin. Average performance would further reduce the percentage.
4. The student would have no clue, even though he had studied Latin four years, to ~~one~~ fourth of the English words derived from Latin because the source words had not been used by the traditionally read authors.
5. It is difficult upon occasion to try to decide why a word, for instance the preposition prae, is not introduced until the third year. In light of its frequency of combination in both Latin and English words, this delay in presentation seems highly impractical.
6. Some English and Latin compounds, parts of which are allocated to different years of Latin, pose as problems; especially when one must wait two or three years to teach a compound, assuming that

Recommendations

In view of the findings and conclusions, the author of this study deems it desirable to suggest certain recommendations in regard to the College Board List. It would seem to be a practical step, since derivation is regarded as an objective by the authors of modern Latin textbooks, for the College Board to recognize the contribution of Latin to English vocabulary insofar as that aim would not interfere with the major purpose of the Latin Word List.

Improvement of the instrument might be achieved if some of the following measures were introduced:

1. Experimentation in the use of new reading material in the first and second years designed to incorporate the suggestions mentioned in numbers 2 - 5 below.
2. Replacement of those words in the College Board List which supply no derivatives, unless they are frequently used in the Latin selections to be read or needed for the teaching of sentence structure and transition.
3. Placement in the first and second year of those Latin words which are rich in derivatives and now assigned to the vocabulary of third and fourth year Latin, and of certain prolific words not included in the College Board List in order that they may occur within the vocabulary experience of the great majority of students of high school Latin.
4. An increase in the number of Latin words which provide our thought vocabulary in English. These words may be referred to as our expressive words and descriptive terms and are learned from the printed page. They may be learned through the study of Latin and may well become the bulk of a Latin Word List used as a fundamental vocabulary.
5. Removal of many Latin words which yield our functional, everyday common knowledge vocabulary which is acquired through imitation and vicarious experience, unless such words satisfy the major purpose of the study of Latin.
6. A new statement of the objectives of high school Latin with a view to bringing them into close harmony with the general objectives of secondary education.

the student takes advanced Latin. For instance the word, "bilingual" comes from bis, a first year word, plus lingua, a fourth year word in the College Board List. If the sophomore in high school meets the new word "bilingual" in his reading, he may know bis means "twice" but he has no key as yet to lingua; and he stops at the end of two years, as he more than likely will, he will never have a key through the Latin vocabulary of the College Board List to the meaning of this word and numerous other such words.

7. Many of the words postponed to the third and fourth year vocabulary are identical with or so similar to common English words that the first year Latin or the non-Latin student could determine their meaning—such words as axis, duplex, templum, monstrum, solidus, liquidus, herba.
8. Words which are identical in Latin and English, like honor, labor, ego, minor, which fall into the vocabulary of first and second year Latin, are usually too obvious to need special emphasis. Although their meaning can be gleaned from vicarious experience, some of them make interesting word studies; for example, the change in meaning of sinister, and the duty of a pastor.
9. Many important words not included in the College Board List are included in derivative word studies by authors of textbooks, but their effectiveness is greatly impaired because they are not contained in the vocabulary set for mastery.
10. One may assume that a knowledge of English prefixes derived from Latin prepositions is very important in developing understanding of word formation and meaning. Latin prepositions were used most frequently of all of the words in the College Board List in building English words in the sample.
11. It is significant that one-half of the words in the College Board List which did yield words in the sample were used only once. Even more convincing is the fact that a certain seven words contributed almost as many derivatives as did 464 other Latin words.
12. Alteration and/or implementation of the College Board List would mean a change in the reading material of the first two years of high school Latin inasmuch as the Latin Word List is the compilation of the vocabulary accepted as the norm for reading.

# A CRITICISM OF STUDIES UTILIZING THE "INDEX OF INDUSTRIOUSNESS" AS A MEASURE OF WORK HABITS

ARVO E. JUOLA

University of California  
Berkeley, California

A SERIES OF articles (1, 2, 3, 4, 5) has recently been published in several educational journals by W. C. Krathwohl in which he stresses the value of what he calls the "Index of Industriousness". He asserts that this scale provides a measure of industriousness or work habits which enables the individual to achieve over and above his ability level as revealed by an aptitude test in a particular field. Furthermore, he points out that when ability level is held constant this scale has value in discriminating the more successful students from the less successful. An additional indication of the value of the I. I. (Index of Industriousness) was cited when it was found that all of the students in English and mathematics who scored low on both the aptitude test and on the I. I. scale in each area dropped out of school within two years.

From these data it would seem that the I. I. is of definite value and applicable to many counseling situations, for as Krathwohl states (5): "The great advantage of the present theory is the ease with which a counselor can secure information concerning the work habits of a student from his test scores."

## Definitions

The use of this technique necessitates a separate I. I. for each subject field which is to be considered. The I. I. for mathematics, for example, consists of a student's derived score on a mathematics achievement test minus his derived score on a mathematics aptitude test. The derived scores consist of a linear transformation from which a mean of 20 and a standard deviation of 4 is obtained for each test.

To interpret these scores students who are in the top 25% of the I. I. distribution are considered industrious or achieving above their measured ability level. The middle 50% are called normal or achieving according to expectations, while the bottom 25% are designated as indolent or achieving below their abilities.

## Findings

In each of his articles, Krathwohl attempts to show that his scale is useful in that for a given degree of brightness as measured by his aptitude test scores, the harder the student works or the higher his score

on the I. I. scale, the greater will be his achievement.

An attempt was made to validate this hypothesis by citing the correlation coefficients between grades in college algebra and scores on the I. I. scale when only students who received identical aptitude test scores were compared (2, 5). Table I contains these correlations. They would seem to indicate that the I. I. bears a rather high relationship to algebra grades over and above that which is accounted for by the aptitude test.

Aptitude scores were also combined into categories which were called above average, average, and below average. The distributions of grades were then compared with the I. I.'s for each of these levels in a manner similar to that which had been done in considering the distributions for each aptitude score separately (3, 4).

Furthermore, the grade means obtained by students who had been rated on the I. I. scale as industrious, normal, or indolent for the same levels of brightness were compared (1). The means were always markedly different. When the levels of brightness were neglected, however, as when the I. I. scores were correlated with English grades without considering each aptitude level separately the correlations dropped to .06 (3).

### Criticism

Krathwohl states that when the groups are taken as a whole, the I. I. contributes little to achievement in the subject field. His major premise seems to be to hold aptitude constant, and then to study the influence of I. I. for equal levels of ability.

Let us, however, consider the original definition of his Index of Industriousness. It was said to consist of the achievement test derived score minus the aptitude test derived score. If each group of I. I. scores is to apply only to equal levels of aptitude, this means that all students who have obtained a given aptitude score received their I. I. score by having their achievement test scores reduced by the same constant. In Table I, for example, the 84 students who received an aptitude test score of 27 earned their I. I. score by having 27 subtracted from each of their achievement test scores. The distribution of I. I.'s at each of these levels for aptitude no longer refers to unique scores for industriousness. They have, instead, been reverted back to a distribution of achievement test scores for each level of aptitude. Each of these achievement test scores has been reduced by a constant but this does not alter the distribution for comparative purposes.

It is, therefore, apparent that Krathwohl has converted aptitude and achievement test scores into his I. I. He then says that this I. I. is of little value for predicting grades if applied to the entire distribution without regard to levels of aptitude. This I. I., then, seems materially to aid

TABLE I  
CORRELATION COEFFICIENTS BETWEEN GRADES IN COLLEGE ALGEBRA AND INDEXES  
OF INDUSTRIALNESS FOR MATHEMATICS

N	Mathematics Aptitude	r	Confidence Level	N	Mathematics Aptitude	r	Confidence Level
84	27	0.49	1%	125	21	0.42	1%
76	26	0.45	1%	128	20	0.35	1%
121	25	0.51	1%	85	19	0.30	1%
157	24	0.48	1%	65	18	0.27	3%
186	23	0.49	1%	65	17	0.33	1%
163	22	0.42	1%				

prediction only when that special case is taken in which the I. I. is transformed back to achievement test scores which he had in the first place. The correlations in Table I which are listed for each level of aptitude are no more than correlations between the achievement test scores and grades in college algebra.

The same reasoning applies in comparison of the grade means of the students with the same degree of brightness but who had been designated as industrious, normal, or indolent by their scores on the I. I. The industrious group consisted of no more than the highest 25% on the achievement test since the I. I. had been reverted back to the achievement test scores when only specific aptitude levels were considered. The industrious, normal, and indolent groups which he uses for prediction could have been equally as well selected from the original distribution of achievement test scores.

The reversion of the I. I. back into achievement test scores is not as apparent when the level of brightness is held constant by the grouping of aptitude test scores into above average, average and below average groups (3). The reduction of achievement test scores by a constant no longer directly applies. It is difficult to define exactly the transformation in achievement scores that has occurred since the data in the publications are incomplete. It seems, however, that instead of a definite constant being subtracted from each achievement score, the I. I. scores tend to approximate the achievement scores minus the mean aptitude scores for each of the groups. The I. I.'s misleadingly appear to fit the original designation as being the difference between the aptitude test scores and the achievement test scores. Actually, however, since aptitude was held relatively constant, the resulting I. I. scores have become more nearly like the original achievement test scores. It is only by enhancing this resemblance that the I. I.'s are made to bear a high relationship to class grades.

It was found that decreased correlation coefficients between the I. I. and later achievement were apparent when a wider range of aptitude scores was used to define groups (3). This finding relates directly to the latter view in that the more unlike the achievement test scores the I. I.'s become, the lower is their relationship to class grades or later achievement.

### Summary

Careful consideration of the available facts seems to justify the following conclusions:

1. The Index of Industriousness is an interesting and perhaps illuminating way of measuring the influence of a student's work habits,

but its relationship to class grades in English cannot be considered higher than was given by the correlation coefficient of .06. Since the relationships to other subjects were not given, they cannot be inferred to be any higher.

2. Considering the relationship of the I. I. to later class grades for only specific aptitude levels is meaningless. The I. I. when this is done loses its meaningfulness and is reverted back to the original achievement test scores with the aptitude held constant.
3. A much simpler and more practical procedure for predicting class grades is to use the aptitude test and achievement test scores directly.
4. A theory of work habits which is based upon the I. I. seems to have a very weak foundation indeed, since the only condition under which the I. I. can be considered useable lies in reverting the scores back to the measures from which they were obtained.

#### KRATHWOHL REFERENCES

1. Krathwohl, William C. "The Persistence in College of Industrious and Indolent Work Habits," Journal of Educational Research, XLII (January 1949), 365-370.
2. \_\_\_\_\_ "Effects of Industrious and Indolent Work Habits on Grade Prediction in College Mathematics," Journal of Educational Research, XLIII (September 1949), 32-40.
3. \_\_\_\_\_ "An Index of Industriousness for English," Journal of Educational Psychology, XL (December 1949), 469-481.
4. \_\_\_\_\_ "Relative Contributions of Vocabulary and an Index of Industriousness for English to Achievement in English," Journal of Educational Psychology, XLII (February 1952), 97-104.
5. \_\_\_\_\_ "A Theory of Work Habits of Industriousness," Journal of Engineering Education, XLII (November 1951), 157-163.

Y



## BOOK REVIEWS

L. Joseph Lins, University of Wisconsin; Valworth Plum, University of Minnesota;  
John Schmid, Michigan State College

Humphrey, George. Thinking (New York: John Wiley and Sons, Inc., 1951),  
pp. ix + 331, \$4.50.

This scholarly introduction to the experimental psychology of thinking was written by the Director of the Institute of Experimental Psychology, University of Oxford. The author is thoroughly at home in this field for he has written on special phases of the topic. His Directed Thinking (1948) is one of the best of the more popular treatments.

The chapters are organized around historical trends and controversial issues. The first chapter, for example, reviews associationism in its many forms; it is stated, in summary, that the psychology of thought consists largely of an unsuccessful revolt against this doctrine.

A hundred pages (about one-third of the book) are given to the work of a group of psychologists at Wurzburg around the beginning of this century. The controversy over imageless thought and the development of such concepts as the determining tendency are related from a careful study of the extensive German publications. These chapters will constitute, undoubtedly, the standard reference to the Wurzburg School, for, probably, no one will attempt this job again with such scrupulous attention to the details of translation and exposition.

Selz' continuation and modification of the Wurzburg approach—not elsewhere available in English—is described and also the Gestalt point of view. A chapter is devoted to thought and motor reaction and another to language and thought. Finally, there is a chapter on generalization which includes such topics as methods of studying concepts, the function of imagery in generalization, and the conditions of efficient generalization.

All of these topics are clearly set forth. Pertinent evidence and arguments from both sides of the Atlantic are surveyed. Most chapters end with a critique and helpful summary. The concluding chapter brings together sixteen general statements about thinking.

Humphrey was not writing a systematic account of thinking. He was not directly concerned with educational practice. This is a critical review of those aspects of thinking which psychologists have chosen to investigate. Anyone who wishes a cogent introduction to what has been done, with shrewd evaluations of the results, and anyone who wishes to plan further investigations would do well to begin with this book.

Donald M. Johnson

Michigan State College

## BRIEF COMMENTARIES

Havighurst, Robert J. Human Development and Education (New York: Longmans, Green and Co., 1953), pp. ix + 338, \$4.00.

What skills or tasks must be mastered by a child as he grows; at what age does the average child master these tasks? In answering these questions, Havighurst has developed a very useful reference book for persons interested in child and adolescent development or, in fact, development throughout the life cycle. He stresses objectives of education and evaluation of educational outcomes.

L. J. L.

Hymes, James L., Jr. Effective Home-School Relations (New York: Prentice-Hall, 1953), pp. vii + 264, \$3.50.

Hymes presents, in an interesting and easily read style, both the theory and practice of improving relationships between teachers, parents, and the community. Though it would appear that adult programs are beneficial primarily to the parent and the teacher, Hymes points out that the real gain is to the child, for only in a continuously congenial, well coordinated and well integrated environment can the child do his best. Hymes has lectured to parent-teacher groups for some time; his book attests to that experience.

L. J. L.

Lawson, Douglas E. School Administration Procedures and Policies (New York: Odyssey Press, 1953), pp. ix + 405, \$3.50.

Though apparently written for a basic course in school administration, the experienced administrator well might read this text. Lawson attempts to present a functional type of administration with solutions to problems being as carefully worked out and arranged in a "step-by-step" form as a manual for some laboratory course.

There is no question that this book can be very helpful to the school administrator in pointing the way to a solution to a problem; however, one must keep in mind that the solution presented may not be the only, or even the best solution, though Lawson seems to imply, in some instances, infallibility of technique.

L. J. L.

Lee, Gordon C. An Introduction to Education in Modern America (New York: Henry Holt and Co., 1953), pp. xiv + 555, \$4.50.

This book presents a survey of American education including American traditions, the experimental approach, organization and administration of American education, the preparation and selection of teachers, social forces, and contemporary problems. Each chapter is followed by questions intended for discussion supplemental to the presentation in the text.

L. J. L.

Simpson, Ray H. Improving Teaching-Learning Processes (New York: Longmans, Green and Co., 1953), pp. ix + 487, \$5.00.

Two needs were considered in writing this book: (1) the need of help to teachers in the junior and senior high schools in meeting teaching-learning situations, and (2) the need of bringing practical applications to prospective teachers through college methods courses.

L. J. L.

Strang, Ruth. The Role of the Teacher in Personnel Work (New York: Bureau of Publications, Teachers College, Columbia University, 1953), pp. xvi + 491, \$3.75.

This is the fourth revision of The Role of the Teacher in Personnel Work, the other editions being 1932, 1935, and 1946. Each edition has had presented, to greater degree than previous editions, student personnel work with concrete illustrations. The book is not divided according to elementary, secondary, and higher education, but rather continuity of guidance throughout the educational experience is emphasized.

L. J. L.

Talucci, D. A. (Ed.) The Punched Card Annual, Vol. II (Detroit, Mich.: The Punched Card Publishing Co., 1953), pp. 208, \$6.00.

The Punched Card is a splendid collection of reports designed primarily as a reference manual of applications of punched cards to business and industry. Some of the reports would be useful to educational institutions in business operations but the manual is of very little value to the research worker.

L. J. L.

## Research News and Communications

Lawrence P. Blum, University of Wisconsin Extension, Milwaukee; Jacob O. Bach, Southern Illinois University; Lester Sands, Santa Barbara College

Study Development of Adult Education in America. The hypothesis, which is being explored at Florida State University, is that there are certain persistent educational needs and developmental tasks of adults in American life which society has attempted to meet by establishing organized programs of adult education. This exploratory study seeks to determine from the literature of adult education what are these persistent educational needs and developmental tasks and to evaluate the effectiveness with which past and current programs of adult education have met them. Indications thus far are that a master chart can be prepared showing some of the organizational patterns and the needs which they attempted to satisfy. Areas of need which were unsatisfied or inadequately met can also be exposed in this manner.

Additional information concerning this study can be secured from Jane E. Guinnane, Coordinator of Undergraduate Counseling, School of Education, Florida State University, Tallahassee, Florida.

A study and analysis of all decisions handed down by the United States Supreme Court which have concerned education has been completed. This study involved scanning all the 343 volumes of Supreme Court decisions, as well as reading many briefs and decisions of lower courts effecting education.

All the cases have been classified under suitable headings in terms of their implications for education.

Further information concerning this study or the data can be secured from William V. Badger, Doctoral Candidate, Florida State University, Tallahassee, Florida.

Journal of Research in Music Education Appears. The first issue of this new publication appeared in the spring of 1953. It is a publication of the Music Educators National Conference which will summarize research studies in the field of music education. It is available by writing to the Music Educators National Conference, 64 East Jackson Blvd., Chicago, 4, Illinois. The single copy price is \$2.00.

Educational Research featured in Phi Delta Kappan. The October 1953 issue of the Phi Delta Kappan is devoted entirely to educational research. This issue was assembled largely through the efforts of the Commission on Research of Phi Delta Kappa. The content developed a broad point of view about research itself and of its contribution to the expansion of education in our society.

Begin selection program for business schools. A group of schools of business administration concerned about the admission of graduate students have been working toward the development of a selection testing program. A conference was held at Princeton this spring, and the conference recommendations subsequently discussed at the annual meeting of the American Association of Collegiate Schools of Business. The general intent is to begin with a testing program to measure aptitude for scholastic work in business, and a research program to develop measures of non-academic characteristics required for success in school and business. As valid measures of the latter become available, they will be incorporated in the selection testing. A number of schools have already agreed to support such a program, and plans for it are now being formulated. Additional information can be secured from the Educational Testing Service, Princeton, New Jersey, which is cooperating on this study.

Study factors relating to failure and success in engineering studies. Graduation and withdrawal rates among fall of 1948 freshman entrants to 101 engineering colleges were recently studied for the U. S. Coast Guard. By July of 1952, one-third of the entrants had graduated or had completed four years of a five-year program; 11% were still enrolled; and 56% had withdrawn of their own volition or been dropped.

Of the students who withdrew voluntarily, nearly half were failing, but a substantial proportion (28%) withdrew in good academic standing in order to change their curriculum. Graduation rates varied by geographic region and by type of institution; for individual institutions they ranged from 12% to 68%. Further information can be secured from A. Pemberton Johnson, Educational Testing Service, who assisted in conducting the study.

The 1954 Conference of the Association for Supervision and Curriculum Development will be held in Los Angeles, March 7-12. Although it had been announced earlier that San Francisco would be the convention city, difficulties in housing convention delegates necessitated the change. The annual conference will be held a month later this year to avoid conflict with second semester registration in colleges and universities.

The Eighteenth Education Conference, sponsored jointly by the Educational Records Bureau and the American Council on Education, was scheduled for the Roosevelt Hotel, New York City, on October 29 and 30. The theme of the conference was "Strengthening Education—All Levels."

Speakers included: Harry J. Carman, Columbia College; Harry D. Gideonse, Brooklyn College; Helen Blair Sullivan, Boston University; Paul L. Dressel, Michigan State College; Jacob S. Orleans, City College of New York; Ralph Steele, JCET; Robert T. Rock, Fordham University; Robert L. Thorndike, Columbia University; and C. Winfield Scott, New Haven State Teachers College.

Study relationship of personality to success in sciences. In the last few years the country has become aware that the supply of outstanding talent is limited and a number of agencies have been set up to inquire into and adjust

the selection and training resources. Psychologists have long realized that some of the differences in high level performance, in research, administration and teaching are no longer determined by differences of ability, but are associated significantly with differences in personality.

This is a first comprehensive research, dealing with a whole range of personality dimensions, planning to get definite information on personality differences that may exist between successful persons in various sciences and between research workers, teachers and administrators in these fields. Such a study will be of great value in leading to the development of methods of determining the extent of available talent and in giving sound advice to students of high general intelligence as to which branch of science and which type of application would give greatest scope to their potentialities. Information concerning this study and availability of results can be secured from R. B. Cattell, Research Professor of Psychology, University of Illinois.

## Research Abstracts and Bibliographies

T. A. Lamke, Iowa State Teachers College, Cedar Falls, Iowa

Action. (Detroit: Metropolitan Detroit Bureau of Cooperative School Studies, 60 Farnsworth St.), 10¢ per copy.

The organ of the Bureau of Cooperative School Studies, this publication appears periodically. Vol. VI, No. 3, April 1953, deals with the subject "Making the Lunch Hour an Educational Experience," and contains a 15 item bibliography.

Adam, R. S. The Relationship Between Age and Reading Attainment in Fiji Schools. (Suva, Fiji: Department of Education, 1953), 23 pp., 2 s. 6 d.

A study of the non-European children in the Fiji Islands. Results suggest that for these pupils the optimum age for the introduction of English reading is 7 or 8 years. Children beginning too early make poor progress; children beginning later are handicapped, though they may recover lost ground later.

Anderson, Kenneth E. A Summary Report to the North Central High Schools of Kansas on Criterion 5: School Plant and Equipment. Kansas Studies in Education, Vol. 3, No. 3, April 1953. (Lawrence, Kansas: University of Kansas, School of Education, 1953), 24 pp., n.p.

A status study of the physical facilities of high schools accredited by the NCA. A supplement on "Acoustical Considerations of School Buildings," by James F. Nickerson, contains a 10 item bibliography.

Annual Report of the Federal Security Agency, 1952. (Washington, D.C.: U. S. Government Printing Office, 1953), 31 pp., 15¢.

Includes a list of publications issued by the U. S. Office of Education for the fiscal year 1952.

Barker, Roger G., and others. Adjustment to Physical Handicap and Illness: A Survey of the Social Psychology of Physique and Disability. Social Science Research Council Bulletin 55, Revised. (New York: The Council, 230 Park Avenue, 1953), 440 pp., \$2.00

A critical review and appraisal of several hundred studies on various aspects of adjustment of physical handicap and illness. Contains a bibliography of 921 items.

Beuschlein, Muriel and James M. Sanders. "Free and Inexpensive Teaching Materials for Science Education." Chicago Schools Journal Supplement, Vol. XXXIV, Nos. 5-6, Jan.-Feb., 1953. (Chicago: Chicago Schools Journal, 6800 Stewart Ave., 1953), 48 pp., 10¢

A classified list checked for current availability and soundness of materials from an educational standpoint.

Davis, Annie Lee. Children Living in Their Own Homes. Children's Bureau Publications No. 339—1953. (Washington, D.C.: U.S. Government Printing Office, 1953), 52 pp., 20¢

Sets forth the range of social services that should be available in each community through child welfare programs to help parents in their task of child rearing. Children living in their own homes who may be in need of social services are identified. The variety of social services required is described.

Education in Great Britain. (New York: British Information Services Reference Division, 30 Rockefeller Plaza, 1952), 47 pp., n.p.

An outline of the history and present status of education in Britain. Contains a 20 item bibliography.

Frank, Lawrence K., and others. Personality Development in Adolescent Girls. Monographs of the Society for Research in Child Development, Vol. XVI, Serial No. 53, 1951. (New Orleans, La.: Child Development Publications, School of Medicine, Louisiana State University, 1953), 316 pp., n.p.

Projective techniques were utilized to explore the personalities and emotional reactions of 300 girls at the period of puberty, with the results here reported.

Fruthey, Fred P. Differential Characteristics of the More Effective and Less Effective Teachers. (Washington, D.C.: Extension Service, U.S. Department of Agriculture, February, 1953), 14 pp., n.p.

A summary report of nine studies made for the Office of Naval Research. Eight of the nine studies were conducted in the field of informal adult education while one study was conducted with children in formal school situations.

Gates, Arthur I. Teaching Reading. What Research Says to the Teacher, No. 1. (Washington, D.C.: NEA, 1201 16th St., N.W., 1953), 33 pp., 25¢

The first in a projected series of pamphlets designed to help the classroom teacher keep pace with the continually advancing field of educational research. Contains 33 selected references.

Guide for Health Counselors. Curriculum Bulletin 1952-53 Series, No. 3. (New York: Board of Education of the City of New York, 1953), 105 pp., n.p.

Designed for secondary school counselors, the Guide outlines approved methods and techniques for health screening, counseling, follow-ups of health problems, and provision of health services in case of sudden illness or accident. Contains a 73 item bibliography.

Horn, Francis H., editor. Current Issues in Higher Education 1953. (Washington, D.C.: NEA, 1201 16th St., N.W., 1953), 292 pp., \$3.00

Outlines proceedings of the 8th annual National Conference on Higher Education held in Chicago in March 1953. Contains addresses and summaries of discussions dealing with all phases of higher education.

# JOURNAL of EDUCATIONAL RESEARCH

Volume XLVII

January, 1954

478273  
Number 5

## RESEARCH ON SCHOOL BOARD PERSONNEL: CRITIQUE AND PROSPECTUS

W. W. CHARTERS, JR.  
University of Illinois

DURING THE past quarter of a century, over one hundred empirical studies of the social characteristics, attitudes, and qualifications of school board members have appeared in the educational literature.<sup>1</sup> Many of the studies use identical research procedures to collect similar kinds of information about a variety of school boards across the United States and, in this sense, constitute a series of replicated studies on a scale rarely achieved in educational research. The time is appropriate to review and evaluate the contributions of these studies to a science of education. In this paper, we will be particularly interested in discovering what knowledge of lasting value has emerged from this mass of research. We will ask the question: What empirically grounded conclusions concerning school board personnel exist today which did not exist twenty-five years ago?

For the convenience of discussion, we have classified the studies into two categories according to the use they make of their data: status studies, which describe the "social characteristics" or other attributes of school board members at a particular place and at a definite point in time, and analytical studies, which attempt to show empirical relationships between two sets of variables regarding board members. The analytical studies mainly are devoted to discovering the kind of person who is best qualified to serve on a school board, they proceed by establishing relationships between criteria of board member competency and the various "social characteristics" of the board members. These studies differ from one another primarily with respect to the variables they use as criteria of competency.

Certain of the status studies share with the analytical studies an interest in the relationship between a board member's competency and his social characteristics. In these status studies, the investigator's attempt to evaluate the general quality of board members in the territorial unit for which they have collected data, based upon assumptions

1. A high proportion of the studies are unpublished dissertations. The author will gladly furnish a copy of his bibliography upon request.

regarding attributes which indicate their competency. Such studies do not belong in the analytical category, however, because they simply assume relationships, while the analytical studies attempt to establish these relationships empirically.

### 1. Status Studies

In examining the empirical studies of school board members, it is unusual to find one that is not a status study. More than 75 surveys of the "social characteristics" of board members have been reported since the earliest appeared in 1904, many of which seem to have been inspired by George Count's classic monograph of 1927, The Social Composition of Boards of Education.<sup>2</sup>

The research operations in all of these studies are similar: the investigator compiles information regarding the age, occupation, income, sex, and other standard social characteristics of school board members in a particular territorial unit. He may also collect information concerning recreation and reading habits, past political offices, teaching experience, religion, civic club membership, length of service on the board, and even the attitudes of board members. Information typically is obtained by mailed questionnaires from school superintendents or from board members directly. The investigator tabulates the returns, presenting the data in the form of averages or frequency distributions.

Studies differ primarily with respect to the territorial unit or type of school district whose members are canvassed. The unit may be the United States as a whole or, more commonly, smaller units such as Pennsylvania's third and fourth class districts or the independent city districts of Georgia. A few investigators restrict their surveys to a single school board, compiling information for the various members who have served on it over a period of time.

### Statistics from the Status Studies

From these studies we discover facts of the following order: we learn that 67.1 percent of board members in Oakland County, Michigan, in 1934, had no training beyond the elementary school, while 45 percent of the Oregon rural school board members in 1937 and 27 percent of the central rural school boards of New York State in 1940 had not gone beyond the eighth grade in school. The proportion of women on various kinds of school boards at various times runs from 3 to 20 percent; the

2. Supplementary Educational Monographs, No. 33 (July, 1927.) Four additional status studies are reported to be in process by the Midwest Administration Center's "Bibliography of Research and Publications Concerning the Functions of School Boards," May, 1952

proportion of board members with children in school may vary from 50 to 90 percent. It would be possible to summarize in detail all of the findings of the numerous status studies, insofar as they classify board member attributes in a comparable manner, but let us consider the scientific value of this mass of statistics before drawing up such a summary.<sup>3</sup>

The statistics show, in a rough way, certain aspects of the composition of American school boards during the second and third decade of the twentieth century. The data may prove valuable for historians in future years, either as commentaries on the United States as a whole or on the local regions in which studies have been conducted; their historical promise cannot be estimated by this reviewer. Moreover, such data may furnish ammunition for arguments attempting to demonstrate, for example, that there are "too few" women serving on school boards or there is an "under-representation" of laborers on boards of education, and so forth.

Aside from the service the studies may perform for historical description or for polemical debate, their value to a science of education seems to be seriously limited. The purpose of science is to provide verifiable reasons why differences occur between sets of facts; the procedure of science is to show that the differences between facts may be accounted for by another set of facts which are found to vary accordingly. The individual status studies do not seek to establish relationships between two sets of variables, but one might argue that taking a number of the studies together they lay a solid groundwork of facts, the variations among which other studies could later attempt to explain.

Unfortunately, the status studies do not lend themselves to meaningful comparisons. The differences between the proportion of female board members found from one study to the next, for example, are confounded by the circumstance that the studies were conducted at different times and in different kinds of school districts. Even if they were conducted at the same time, it is virtually impossible to define the crucial ways in which the types of school districts differ. What differences are there between the six-member boards in Nebraska and the boards of Schuylkill County, Pennsylvania, which could explain the difference in the proportion of female members of the two groups of boards? We cannot do more than guess, and an investigator who wishes to account for variations in female membership would find it more efficient to design and execute his own study than to review the statistics which

3. The dissertation by J. W. Owens brings together the findings of a number of status studies in Trends of Thought Concerning Boards of Education as Revealed in Articles Appearing from 1929-1938 in the American School Board Journal, University of Michigan, 1940.

have been amassed over the years.<sup>4</sup>

A few of the status studies verge upon being analytical studies by presenting within the context of a single study the distributions of social characteristics for different classifications of boards or of board members. Counts, for example, showed the distributions of occupations of members of city, county, state, and university boards.<sup>5</sup> Others classify boards according to whether members are elected or appointed, or classify board members according to their length of service, according to whether they are officers or common members, and so forth. In none of these studies, however, are the differences in the distributions of social characteristics among the classifications of major interest to the investigator. In essence, the systems of classification rarely serve an analytical purpose in the research.

### The A Priori Assumption of the Status Studies

A number of the status studies are explicitly devoted to the topic of qualifications for board membership. In contrast to the analytical studies which we will discuss in the following section, they posit a priori assumptions regarding the social characteristics which are most desirable in board members. They assume precisely what the analytical studies intend to establish empirically.

With the assumptions as the point of departure, the research operations in these status studies are directed toward describing the distribution of the social characteristics among a population of school board members. Conclusions are drawn concerning the degree to which the surveyed board members meet the a priori standards. Below are two illustrations of the form of argument in these studies, selected from the writings of Counts and of Bishop.<sup>6</sup>

#### Counts

Assumption: Persons from the "dominant class" are undesirable as board members (at least, when they are in the majority).

4. Moreover, it is exceptionally difficult to ascertain from many of the research reports the extent to which the observed differences are simply consequences of variations in research procedures. Requisite methodological descriptions usually are inadequate and frequently entirely missing from the reports.

5. George S. Counts, op. cit.

6. George S. Counts, ibid., and William E. Bishop, The Social Composition of High School Boards of Education in Colorado, unpublished Master's dissertation, University of Colorado, 1939.

The reason behind this assumption is based upon another assumption, viz., that persons from the "dominant class" are committed to status-quo policies in education.

Empirical Finding:

The great majority of school board members in the United States (in the 1920's) are business or professional men, i. e., members of the "dominant class".

Conclusion:

American school systems are controlled by persons committed to policies maintaining the status-quo.

Bishop

Assumption: Property owners are desirable board members. They are interested in long-time planning for the community, while non-owners are only interested in the present.

Empirical Finding:

The majority of members of Colorado high school boards (in 1939) are property owners.

Conclusion: Colorado high schools are in capable hands (as far as the social characteristic of property ownership is concerned).

It is obvious that empirical research of this sort is not designed to test the validity of the assumptions; thus, no increment in empirically grounded knowledge concerning board membership qualifications emerges from such studies.

Furthermore, the conclusions drawn from these studies can be no more substantial than the assumptions upon which they are based. Most of the investigators attempt to gain stability for their assumptions by citing the opinions of authorities on school administration whose writings preceded their own.<sup>7</sup> Disagreements and contradictions in professional opinion, however, are common with respect to such matters as the value of women or of business men as members of boards of education. With respect to some other matters, such as the value of well-educated board members, professional consensus runs high. But

7. A comprehensive summary of professional opinion concerning the value of various social characteristics of board members may be found in Henry M. Gunn, The Study of a School Board in a Western City, unpublished Doctoral dissertation, Stanford University, 1942.

even consensus of professional opinion fails to provide a stable basis for assumptions as long as the opinion is unvalidated by systematic, empirical research. Authoritative opinion in education, as well as other academic fields, is responsive to the currents of the ideological climate of the profession or of society at large; it frequently reflects in large measure the professional man's own social outlook.<sup>8</sup>

Neither from the standpoint of the accumulated statistics concerning board members nor from the standpoint of an understanding of the qualifications of board members have the status studies contributed substantially to a science of education. The statistics are simply descriptions of unique groups of school boards across the nation--unique in both time and place. When comparisons between the findings of studies appear possible, as in the case of country-wide investigations at two or more points in time, methodological inadequacies of the studies and the absence of correlative data render the comparisons relatively useless.<sup>9</sup> The approach of the status studies to the qualifications of board members has, in effect, put the cart before the horse. Before the quality of board members can be evaluated on the basis of their social characteristics, the relationships between the competency of board members and their various attributes should be firmly established by empirical investigation. We will now turn to the few studies which have attempted to seek these relationships by empirical analysis.

## II Analytical Studies

Virtually every analytical study in the literature of research on school board personnel attempts to answer the question: What kind of a person is best qualified to serve on a school board? The procedure common to these studies involves (a) establishing some criterion by which highly qualified board members can be singled out from among

8. A body of research and theory exists concerning the effect of social conditions on professional opinion and constitutes part of the subject matter of the sociology of knowledge. Cf., Florian Znaniecki, The Social Role of the Man of Knowledge, New York: Columbia University Press, 1940; and Logan Wilson, The Academic Man, New York: Oxford University Press, 1946.

9. Particularly relevant here is the observation that few investigators have shown any regard for the problem of sampling bias. Bias is markedly present when investigators rely upon mailed questionnaire responses for their data--the method upon which all of the country-wide studies depend. By not following systematic sampling procedures, studies conducted of the same population of board members at different points in time cannot be compared with any assurance that differences revealed by the comparisons are true differences and not due to differences in method.

the less qualified members and (b) isolating those social characteristics which distinguish the highly qualified members from the others. From this empirical procedure, investigators propose to describe the well qualified school board member in terms of his occupation, age, sex, and other attributes.

Three criteria have been used in these studies to judge the quality of board members: (1) their voting record on issues confronting the board, (2) ratings by school administrators, and (3) their social attitudes.

#### Voting Records of Board Members

The two investigators who have adopted voting records as their criterion of board member competency reasoned that if board members with certain social characteristics are better qualified for board service than members with other social characteristics, the differences should be reflected in the way the members vote on important issues confronting the board.<sup>10</sup> Consequently, the investigators scanned minutes of board meetings for important issues of various kinds on which the boards had taken action by recorded vote. Ballots cast by each member were judged as representing a "sound" or an "unsound" vote, according to professional standards adopted by the investigators; the proportion of "sound" votes cast by a member constituted the criterion of his competency. By comparing the more competent members with the less competent members, the investigators sought to discover whether or not they differed with respect to their various social characteristics.

Using data from school boards in twelve Western cities, Campbell reported that he found no relationship between competency and such characteristics as age, sex, education, income, occupation, tenure on the board, or parenthood. Gunn drew data from board minutes in a single city (Portland, Oregon) covering a 25-year period and, similarly, found no relationship between the voting records of board members and their various social characteristics. Neither of these studies support the a priori assumptions which other investigators have taken for granted.

We must not be too hasty in concluding that these two studies prove, once and for all, that quality of board service is unrelated to the mem-

10. Both studies were Doctoral dissertations completed at Stanford University in 1942: Ronald F. Campbell, Social Implications of School Board Legislation, and Henry M. Gunn, op. cit.

ber's social characteristics.<sup>11</sup> We must stop to consider how adequately a member's voting record the quality of his service on the school board. In the first place, there is reason to believe that a board member's vote does not universally represent the stand he has taken on an issue. In some school boards, at any rate, members feel obliged to go on record as being in agreement, whatever their stands may have been during the deliberations which preceded the vote. Both Campbell's data and the author's own studies show a remarkably high number of unanimous votes on the issues confronting school boards--a unanimity which misrepresents the true division of opinion among board members.<sup>12</sup> More appropriate data for measuring the stands which board members take during board meetings must come from observing the deliberations of the board prior to the recorded vote.

In the second place, we must consider whether passing on school legislation is the only service or even the most important service which a member performs for the school board. Is it not important, for example, for a board member to serve the school well in the field of public relations? If other services are important for membership qualification, then a criterion of competency based solely on the member's voting record provides an incomplete description of the quality of his service.

11. Both studies contain serious methodological deficiencies which we can do no more than illustrate here. Campbell, in analyzing the relationship between voting and occupation, was only to classify the occupations of board members into two groups-- proprietors and professional men. Virtually all of his 172 board members were what Counts would call representatives of the "dominant" class. Hence, Campbell was unable to compare "dominant" class members with, say, blue collar workers who constitute the "unprivileged" class and thereby to test the assertion of Counts that "dominant" class representatives are less qualified than "unprivileged" class representatives. In addition, Campbell frequently used inappropriate tests of ~~art~~ statistical significance and improperly evaluated those he used. It is impossible to evaluate Gunn's study in detail since he failed to describe the most important aspects of his methods; his presentation of data is too inadequate and confusing to permit the reader to reconstruct his methods of procedure.

12. The author found, in one school board in an agricultural community of Central Illinois, 100 unanimous votes of a total of 107 votes recorded during a two-year period. Two of the seven dissenting votes were simple formalities which occurred when a board member voted against himself for an office. A city school board in Illinois voted on 625 issues in two years; all but 16 of these were unanimous. Again, two of the 16 were formalities. With so few dissenting votes, it is impossible to apply the kind of analysis which Campbell used in his study.

Finally, we must recognize that an element of arbitrariness enters these studies at the point where the investigators make judgments concerning the "soundness" of board member's votes. Campbell judged, for instance, that members who vote in favor of measures to extend the scope of the school program are more satisfactory than those who do not. Gunn assumed that opposition to general financial retrenchment of the school indicates greater competency than its support. Judgments of this order, even when confirmed by professional opinion, represent a point of view which persons in other institutions of the community may not share. For example, where extension of the school program into areas of community health or delinquency control may be favored by educators, others may regard this extension as exceeding the proper function of the school.<sup>13</sup>

#### Administrators' Ratings of Board Members

In the studies of Cooke and his colleagues, each board member was rated on a 21-item list of traits by an administrator in his school system. Those who received high rating scores (the "good" board members) were compared with the remainder, a procedure which enables the investigators to identify the social characteristics which distinguish the "good" members from the rest.<sup>14</sup> A sample of the items on which board members are rated is shown below:

- Item 1. Recognizes superintendent as head of school system.
- Item 2. Has an interest in the schools.
- Item 5. Is intelligent
- Item 12. Is willing to learn.
- Item 14. Is open minded.

13. Moreover, application of abstract standards, such as those illustrated above, to board actions is a questionable procedure when the context of the issues on which the votes are cast is disregarded. During periods of fiscal crisis in the community, financial retrenchment of the schools may well be the most strategic stand for a competent school board member to take.

14. Dennis H. Cooke and Quill E. Cope, "He is a Good Board Member," The Nation's Schools, Vol. 21, No. 6, June, 1938, p. 31-33; Dennis H. Cooke and Con T. Welch, "Portrait of a Good Board Member," Ibid., Vol. 27, No. 2, February, 1941; Quill E. Cope, Personal Characteristics of the Best County School Board Member, unpublished Master's dissertation, George Peabody College for Teachers, 1936; Con T. Welch, Personnel Study of County School Boards of West Tennessee, unpublished Master's dissertation, George Peabody College for Teachers, 1939. For a description of the method used to construct the rating form, see Dennis H. Cooke and Quill E. Cope, "Rating School Board Members," The Nation's Schools, Vol. 21, No. 2, February, 1938, p. 34-36.

From studies of some 230 county school board members in Tennessee, Cooke reports such findings as these: "good" board members are better educated, have higher incomes, own more valuable property, more frequently are proprietors, and more frequently are church and civic club members than board members not classified as "good".

Unfortunately, Cooke's conclusions cannot be accepted at face value since he does not evaluate the statistical significance of the differences between "good" members and the remainder, and some of his conclusions are based upon differences which could easily arise by chance.<sup>15</sup>

Disregarding the methodological shortcomings in these particular studies, let us be clear about the kind of criterion of board member competence used in the general approach. Where Campbell's and Gunn's criterion related to the member's stands on school legislation, Cooke's criterion relates to the degree to which members are liked by administrators. By furnishing administrators with a list of traits on which to rate their board members, Cooke attempts to obtain an objective description of the competency of the members. But rating forms are notoriously susceptible to the "halo effect" (by which raters tend to judge persons they like uniformly high on all items, and to rank those they dislike uniformly low on all items); this is particularly true when the items call for highly subjective judgments regarding a number of traits barely distinguishable from one another, as in the case of Cooke's form. Moreover, superintendents and principals who serve in their capacities at the pleasure of the persons whom they rate cannot be expected to provide the most disinterested and objective judgments of them. In short, we believe the persons Cooke identifies as "good board members" may more correctly be represented as members who are "best liked by administrators".

Certainly, the ability of school board members to gain the respect and cooperation of the administrators whom they hire is one element of their competency, but to select this ability as the sole criterion of quality of board service is to overlook the fact that board members have other responsibilities than that of establishing satisfactory working relationships with administrators.

15. The published reports allow the reader to compute significance levels for some of Cooke's findings but not for all. One of his larger differences, for example, (showing that 12% more "good" members than ordinary members are proprietors and that 13% fewer "good" members than ordinary members are agriculturists) is based upon comparisons of 23 "good" members and 78 members not classified as "good". With a sample this small, a difference of at least 17% is required to meet statistical significance at the .05 level of confidence.

Social Attitudes of Board Members

George Counts believed that proprietors, business and professional men, and other representatives of the "dominant" class are undesirable board members because, as he assumed, their social outlook is conservative. If we accept the belief that one's social attitudes constitute a justifiable criterion of board membership competency, it is possible to discover the kinds of persons who hold progressive, or non-conservative, attitudes. To the author's knowledge, no recent studies have undertaken this type of analysis. Two studies appeared in the early research literature, before the development of reasonably accurate instruments for attitude measurement; it would be of no advantage to devote space to them here.<sup>16</sup>.

III. Research for the Future

In the preceding sections, we have reviewed the major classes of research on school board personnel with the purpose of evaluating their contributions to a science of education. On the whole, the yield is disappointing. The overwhelming majority of research enterprises have been surveys which simply describe the existing state of affairs at a certain time and place. The remaining studies have fixed their attention upon ascertaining the social characteristics of competent board members, but these studies have been too few in number, too weak in methodology, and too restricted in scope to add significant conclusions to a fund of knowledge about school board personnel.

The Criterion Problem: Functions of the School Board

The restricted scope of the studies of competent board members arises, in part, from our incomplete understanding of the functions of school boards. If we are to develop criteria of highly qualified board members, it is necessary to specify carefully and precisely what they are to be qualified for. Recognition that board members fulfill a number of necessary services leads to the possibility that different kinds of persons

16. The more recent of the two studies is Claude E. Arnett's Social Beliefs and Attitudes of American School Board Members, Emporia, Kansas: Emporia Gazette Press, 1932. We have briefly described this study and its severe methodological limitations in "Social Class Analysis and the Control of Public Education", Harvard Educational Review, (in press). Cf., also, George C. Struble, "A Study of School Board Personnel", American School Board Journal, Vol. 65, October, 1922, p. 48-49, 137-138

may be competent to perform the various services--that we may be unable to designate any single person as the competent board member. Thus, a person whose roots are deep in the community, whose breadth of vision rarely transcends the local region in which he resides, may be the kind of person most competent to perform the function of arousing public support for the schools. On the other hand, a person with more cosmopolitan experience and a more encompassing breadth of vision may be the kind of person most competent to examine the large educational problems confronting the local community.<sup>17</sup>

Research on qualifications for board membership is seriously handicapped by the absence of empirical analyses of school board functions. What part does the school board play in the school system and the social life of the community? The literature of education contains numerous estimations of the role which the board should play in order to obtain a higher quality of education in the schools. But a school board is more than an instrument for better education; it is an institutional arrangement deeply imbedded in the social life of the community and of the school. Broad and penetrating sociological research is needed to discover the various vital functions which the board fulfills in maintaining the integrity of the complex social system of which it is a part.<sup>18</sup> Criteria of competency must take into consideration all of the services which board members render the school and the community, not just the one or two which happen to occur to an investigator. To the extent that different board functions require different abilities usually not found in one person, descriptions of the competent board member will have to be supplanted by descriptions of the combination of members constituting the competent school board.

#### The Reference Point for Research: Board Member or Board?

The studies of school board personnel universally center attention in the individual board member. How old is he? Does he typically

17. This illustration is not purely fictitious. It is suggested by Robert K. Merton's investigation of the influence process in the community. Cf. Merton, "Patterns of Influence: A Study of Interpersonal Influence and of Communications Behavior in a Local Community", in Paul F. Lazarsfeld and Frank N. Stanton (eds.), Communications Research, 1948-1949, New York: Harper & Brothers, 1949, p. 180-219.

18. For an exposition of the method of functional analysis in sociology, see Robert K. Merton, Social Theory and Social Structure, Glencoe, Illinois, The Free Press, 1949, Chapter 1, "Manifest and Latent Functions", p. 21-31. Merton especially emphasizes the necessity of discovering the unrecognized and unintended functions of institutional arrangements, such as the school board.

have children in school? What is the occupation of the best board member? Board members are treated as isolated individuals, taken apart from the social matrix in which they work and live. Implicit in considering the individual board member as the reference point for research is the belief that if we accumulate knowledge about the properties of enough members, we will know all we need to know about the boards which control our public school systems. If we wish to understand the nature of the policies adopted by the school board, presumably it is only necessary to sum the attitudes or other properties of the persons who constitute the board and draw our conclusions.

The concentration of research upon the individual board member apart from his social relationships leaves unstudied some of the most critical issues in educational administration and educational sociology, some of which we note below.

1. The school board is presumed to have the greatest power in shaping the character of the school's educational program of any office in the structure of America's educational institution. Hence, the importance of studying the board. "As is the board, so is the school," to paraphrase the opening paragraph of Counts' classic monograph.<sup>19</sup> But is the presumption correct? There is serious need for research describing the distribution of policy-shaping power between the board and the administrative and teaching staffs of the school. Observation of the superintendent-board relationships in school board meetings represents an ideal starting point for studies of the power distribution.

2. Closely related to the foregoing is the subsidiary problem of making clear what we mean by the term "policy". Manuals of school administration tell us that the board should formulate policy and the administrator should execute it, but the difference between a policy decision and an executive decision is not as self-evident as it appears at first sight. For purposes of illustration, let us consider the difficulties involved in distinguishing between policy and executive decisions on the basis of their effects upon guiding the course of the school organization and the educational program. If we define "policy" as those decisions which shape the course of educational affairs, we meet the following difficulties:

- (a) Some decisions which are intended to establish policy have no effect upon the conduct of school affairs and, thus, are not policy.
- (b) The small, day-by-day decisions made by administrators, teachers, and even custodians frequently more effective in determining the course of school affairs than those decisions

<sup>19.</sup> George S. Counts, op. cit., p. 1

which are intended to be policy-forming.

- (c) In any event, it is impossible to distinguish between policy and non-policy decisions until a period of time has elapsed to determine what effects they have had on the school; consequently, an officer cannot know, at the time he makes a decision, whether or not it will turn out to establish a policy.

Similar difficulties are encountered if we base the distinction between policy and executive decisions upon (1) whether or not the decision relates to "broad" educational problems, (2) whether or not the decision entails delegation of responsibility for executing it, or (3) whether the decision applies to a series of similar cases or to one specific case. What, then, is the distinguishing feature of policy decisions? How does a board member or an administrator know whether he is operating on a policy level or an executive level? Until a consistent, unambiguous distinction is drawn between policy and executive decisions, neither the school officer nor the research investigator will find the two terms helpful to him.

3. What becomes of a decision regarding the school program once it has been adopted by the school board? Is the purpose behind the decision lost as it is transmitted to and put into operation by the teacher or other official responsible for its execution? Studies conducted in industrial settings suggest that decisions may be modified beyond recognition as they pass through channels of communication to the point where they are supposed to be put into operation. If this is true in school systems, especially the larger and more intricate systems, it supports the conclusion that the school board has less power than that vested in it by the formal structure of the educational institution.

4. With regard to the operations of the board, studies are needed to show the process by which members reach decisions. Who is most influential in determining the kind of decisions the board reaches or in determining the kind of issues which are raised for decision? What is the nature of the considerations which board members bring to bear as they decide upon issues? To what extent can board decisions be accounted for by the social attitudes which individual members hold as opposed to the other factors which also impinge upon the decision process?

5. When a man is elected or appointed to a school board, he no longer is the same person he was as a mere citizen of the community. His role as a responsible official of an educational institution places his thoughts and behavior in a new context. How does the board member perceive his role? To whom does he believe he is responsible? How does occupancy of the official position modify a board member's attitudes? For instance, is the frame of reference by which he judges social and educational issues thereby broadened? These questions are related to another question concerning which we need empirical infor-

mation: What are the various motivations which lead citizens to seek or accept board membership?

These are a few of the problems which demand study once the point of reference of school board research shifts from the individual board member to the school board and its composite members as they work and live in an intricate system of social relationships. By adopting a broadened research perspective, the close of the next twenty-five years should see a body of knowledge far more significant than that available to us at the end of the first quarter of a century of intensive research.

# RATIONALE FOR REORIENTING EDUCATIONAL RESEARCH \*

JOHN T. AUSTON  
Michigan State College

## Introduction

IT HAS BEEN suggested that perhaps civilization began when primitive man first conceived a new solution to an old problem.<sup>1</sup> No one of course is certain at just what point in the development of the race this "event" took place, but the first attempt at reflective thinking for purposes of problem-solving probably marked mankind's initial venture in research. As one writer has averred: "The problem of research is as old as organized thought."<sup>2</sup>

Although the problem of research has a long history, Kelley points out that:

It is only within modern times that research activity has attained a social importance which is not merely incidental to the older pursuits, religion, philosophy, education, politics, war and commerce.<sup>3</sup> Apparently research is a problem of growth, the essence of which is evolution, in part due to the recent advances in the physical and biological sciences, but perhaps more particularly due to the functional changes in social organization coupled with a corresponding rise in efforts at interpretation. These rather considerable advances have been somewhat dependent upon the refining of methods, techniques and tools of observation, but more particularly dependent on developments in description and theory construction.<sup>4</sup>

## Historical Precedent

The impact of the methods developed in the fundamental sciences on social and educational research has been both contemporary and considerable. Buswell has called attention to the fact that this impact on

\*This article is based in part on a Ph.D. dissertation completed at the University of Denver. For a more thoroughgoing analysis of new conceptual structures in educational research, see Auston, J.T., "Dimensions of Published Speech Research, 1915-1949", (Unpublished Doctor's Dissertation, University of Denver, 1950).

1. Whitney, F.L. The Elements of Research, (New York, Prentice-Hall, Inc., 1942), 1.

2. Ibid., vii

3. Kelley, T.L., Scientific Method, (New York, Macmillan Company, 1932), 1.

4. Lundberg, G.A., Social Research, (New York, Longmans, Green and Company, 1942), vii

educational research was first noted between 1905 and 1915.<sup>5</sup> Good, Barr and Scates indicate that the increasing emphasis upon educational research gained momentum during the first 25 years of the present century.<sup>6</sup> Taber in surveying five fields of original university-sponsored research in education discovered that the earliest dissertations on educational subjects postdated 1890.<sup>7</sup>

The results of this comparatively recent awareness in education of adapting to research the essentials of a scientific system have included the assembly of large masses of quantitative data, the devising of research techniques particularly suited to education, and the development of broad concepts of central importance in studying and directing education.<sup>8</sup> Nor is the recent awareness of the scientific method confined to the field of education alone. A comparable impact resulting in the development of method, technique and research viewpoint has been noted in social research. In this connection Blumer states: "The most notable feature of the contemporary social sciences is the increasing emphasis on research."<sup>9</sup> It is an emphasis which he suggests, has now become the primary concern of the social scientist.

One index of the proportions which social and educational research have assumed is revealed in the fact that only a partial listing of research enterprises indicates that there are upwards of 300 textbooks on research methodology and the scientific method applied to social and educational research, the large majority of which postdate 1922. Whitney in 1942 listed 46 research foundations occupied with social and educational research and 64 annual institutional reports of research in education.<sup>10</sup> To this may be added the long and growing list of professional, educational and government journals, pamphlets, yearbooks and special reports devoted to abridgements of findings of research projects conducted in almost every realm of human concern.

5. McConnell, T.R., Scates, D.E. and Freeman, F.N., "The Conceptual Structure of Educational Research," Supplementary Educational Monographs, 55, University of Chicago Press, May, 1942, v.

6. Good, C.V., Barr, A.S. and Scates, D.E., The Methodology of Education Research (New York, Appleton-Century-Crofts, Inc., 1935), xiii.

7. Taber, G.J. "Critical Analysis of Research Techniques in Five Fields Education as Presented in Doctor's Dissertations," (Unpublished Doctoral Dissertation, The University of Pittsburgh, 1938), 432 pp.

8. McConnell, T.R., Scates, D.E. and Freeman, F.N., Loc. Cit.

9. Young, P.V., Scientific Social Surveys and Research (New York, Prentice-Hall, Inc., 1939), xxvii.

10. Whitney, F.L., Op. Cit., 276, 429.

The growing body of data and the rapid advances in science gave rise early in the century to a felt need for presenting systematic relational analyses of the then-known end-products of research.<sup>11</sup> Most notable among these in the fundamental sciences, of course, has been the Einsteinian system of relativity in the realms of physics and mathematics.

A similar felt need arising from a comparable accumulation of large masses of educational and social data first made its presence felt as early as 1922.<sup>12</sup> Since that time some 6 similar original investigations have appeared in the professional literature of social and educational research.<sup>13</sup> The mention of these analyses of accumulated educational and social data is not an unintentional confusion of two forms of organization. The prior reference to the synthesis of the Einsteinian variety which attempts to reduce a wide variety of factors to a single principle, or a few coherent principles, is not necessarily parallel to compilations of social and educational research-based data, revealing varying degrees of integration, in any but the broadest structural sense.

So that although nothing in the realm of educational research methodology as comprehensive as an Einsteinian synthesis has appeared, there have been recent indications of professional interest in the development of more systematic relational analyses of the value of the known end-products of educational research. As Einstein has proposed a conceptualized description of the structure of the physical universe, so educators and social scientists since the late 1930's have been proposing that descriptions of new conceptual structures for educational research be undertaken.

Until 1940 the only published material available generally in the broad area of the conceptual structure of educational research was restricted to one or two treatments of the subject in professional journals. Holmes records that as far as he has been able to determine only one such article prior to his 1939 analysis appeared.<sup>14</sup> In 1943, however, the Chicago University Press published a supplementary educational monograph on the Conceptual Structure of Educational Research.<sup>15</sup>

11. To cite only one pioneering example, cf., Pearl, R., Modes of Research in Genetics (New York, Macmillan Company, 1915), 250 pp.

12. Altucker, M.M., The Case Method in Education, Unpublished Doctor's Dissertation, The University of California, 1922.

13. Infra., cf., "Procedural Precedent," below.

14. Holmes, H.W., "Whither Research in Education?" Educational Research (American Council on Education Studies, Ser.1, Vol.3, No.10, Nov., 1939) 1 ff.

15. McConnell, T.R., Scates, D.E. and Freeman, F.N., Loc. Cit.

Other scattered items have appeared which have given impetus to an inclination in this direction. This interest has been chaotic, but nonetheless present in several areas of socio-educational research. As early as 1910 Haret was concerned with organizing the empirically derived facts of sociology into some sort of equation. Subsequent early interest was manifest in the studies of society by Pareto and Sorokin, and more recently by S. C. Dodd and N. Rashevsky,<sup>16</sup> who have speculatively advanced mathematical theories of human relations.

Early in 1947 the American Educational Research Association in response to a special request, commissioned the Florida University Research Bureau to conduct a thorough-going analysis of needed research in education, and to submit recommendations.<sup>17</sup> Rashevsky and others in the field of social research have been concerned with a re-examination of methods and field in sociology in the light of the methodological progress evidenced in the fundamental sciences, for purposes of erecting a comparable value system for social research. More specifically they have been concerned with the deriving of mathematical relationships for diverse sociological quantities.<sup>18</sup> The program proposed by the AERA was less specific than that of Rashevsky and others, but urged adequate support for a five-fold comprehensive and long-range series of educational research projects for purposes of erecting a value system for educational research which would focus on the methodologies and techniques of that research.<sup>19</sup>

The purpose of the present discussion then, is to review the line of reasoning which offers direct support for the implications present in the work of Rashevsky and others, as well as the recommendations of the AERA, both of which have aimed at erecting value systems with which to re-examine the methodology and findings in their respective fields. The broad end-result of applying the rationale could well be the revealing of present status, possible developments and needed theoretical reformulations in educational research.

16. Rashevsky, N., Mathematical Theory of Human Relations, (Bloomington, Indiana, The Principia Press, Inc., 1947), iii-xii (An interesting, brief survey of the efforts of sociologists in the direction of organizing research around a quantified system in light of the known end-products of that research.)

17. Editor, School and Society (February 28, 1948), 160 ff.

18. Rashevsky, N., Loc. Cit.

19. School and Society. Loc. Cit.

Testimony of Experts

In a limited sense what is being suggested here would constitute a probing, on a systematic basis, of the more crucial aspects of the broad conceptual structures of socio-educational research and the essentials of a scientific system, followed by a rigorous application of emerging principles to current research practices. In one sense then such a proposal conforms to the definition offered by Thomas H. Morgan who has indicated that a "...crucial experiment is a study of an event so controlled as to give a definite and measurable answer to a question—an answer in terms of specific theoretical ideas, or better still, an answer in terms of better understood relations.<sup>20</sup>

The rationale for the suggestions incorporated in this paper has been well expressed by G. T. Buswell of the University of Chicago:

"The mass of data derived from research studies now needs to be interrelated and reinterpreted in order to refine the concepts which they support and to integrate these concepts into a coherent and educational theory. More significant hypotheses can then be formulated for the guidance of future research and concepts of still wider application can be derived. This emergence from a level of fact-finding in research to a level of explanatory principles demands some important changes in the structure of educational research. Research has now matured to the point where mere fact-finding is not a sufficient outcome; rather the outcome must now be some explanatory theory supported by the facts already derived through research."<sup>21</sup>

T. R. McConnell in speaking of the importance of outlining systematically the controlling concepts and basic assumptions which serve to focus research on the central problems in major educational areas has pointed out that:

... "a very important reason for the lagging attack on significant matters is a lack of systematic statement of the educational problems which the main fields of educational inquiry present. An outline of central problems in each field should be constructed periodically by taking advantage of the evidence already available, by making a careful survey of the important practical situations, by examining the problems and the data which emerge from investigation and experience in related disciplines, by summarizing all these sources in the form of gen-

20. Morgan, T.H., "The Relation of Biology to Physics," Science (45:1679) March 4, 1927), 217.

21. McConnell, T.R., Scates, D.E. and Freeman, F.N., Loc. Cit., v.

eral principles, and by formulating a set of hypotheses for further scientific investigation. <sup>22</sup>

Isolated though this type of research may be, that there is embryonic recognition of the need for the formulation of new frames of reference capable of reproducing an enormous number of subjects for educational research, each a thrust at the heart of the matter, is suggested in at least two studies recently completed at the University of Minnesota.<sup>23</sup> However the paucity of such designs and approaches is indicated by Holmes who, upon examination of extant reports of educational research concludes that it is:

... "plain we have not discovered any highly developed conceptual structure in educational research. Our study constitutes an invitation to others to dig out of reported research in education a more stately mansion of ideas and principles than we have unearthed; it is also and more significantly a challenge to educational research to state and organize its concepts - to develop a conceptual structure which will give to education a new and greatly needed instrument of progress." <sup>24</sup>

A growing recognition for needed research based on the "field" theory is attested by Watson, who has indicated that research be designed so that mutual relationships and the interdependence of essential factors can be studied, rather than the independent action of a single variable or narrow function.<sup>25</sup> P. R. Mort has compared educational research to a city in the process of development but as yet without a

22. Ibid., 9

23. Cook, W. W., "Grouping and Promotion in the Elementary School," and Horn, E., "Methods of Instruction in the Social Studies," University of Minnesota, Series of Studies in Education, Nos. 1 and 2 (1941).

24. Holmes, H. W., Op. Cit., 3

25. Watson, G., "Research in Guidance and Personality Development," Educational Research (American Council on Education Studies, Ser. 1, Vol. 3, No. 10, 1939), 118 ("field" here refers to the overview approach; Watson's implication is toward both conceptual structure and relational analysis as they bear on research method and emphasis in a given "field" of study) see Austin, J.T.

plan: houses have gone up here and there, public buildings in certain places, industrial plants in other, but as yet there has been no design for the city.<sup>26</sup>

Douglas Scates in a brilliant passage has brought the problem into further focus by calling attention to the over-emphasis on the development of instruments which has appeared as the end-product of most educational research in recent years.

"Is it not somewhat curious that the basic instruments were not forthcoming until the research concepts were well enough advanced to require them? Persons who are prone to explain the history of science in terms of the appearance of instruments alone have not made clear how science would have profited if a cyclotron had been set down in an eleventh century monastery, nor what the Alexandrian school would have done with an X-ray, an electroscope or an infrared microscope. Ptolemy would not have been better off had he had the giant reflecting telescope at Mount Palomar. He would not have known how to operate its intricate mechanisms. Even if we grant him this understanding, he would have spent years in relatively aimless observations; for the great telescope has and will reveal so much that it would be stupefying unless the observer has at his command well developed systems of orienting concepts and highly refined questions backed by a rich foundation of ordered insight. . . . Would one say as in the case of the instruments, that facts are of great importance, but are nevertheless subsidiary to the framework of thought which incorporates them? To this I must reply unequivocally in the affirmative."<sup>27</sup>

In this connection, Holmes makes the following observation:

... "the problem of educational research cannot be solved by a few well-known tricks of method. . . And this means becoming more inclusive in our views, more fruitful of conceptions and hypotheses, more determined in our search for assumptions, more discriminating in our choice of methods, and more imaginative in our whole attack. If there are to be national fellowships for educational research they had better not be given to routine workers."<sup>28</sup>

Alan Gregg in describing the structure of medical research adds the following footnote to the current discussion by insisting that the

26. Holmes, H.W., Op. Cit., 9

27. McConnell, T.R., Scates, D.E., and Freeman, F.N., Op. Cit., 174

28. Holmes, H.W., "Suggestions Toward a Frame of Reference," Op. Cit., 174

"... mere recording of observations and reading of dials on complicated instruments of precision is not research." 29

"What is needed then is an examination of the nature of new facts that must be taken into account, new hypotheses that may be formulated, new issues to be resolved; and thereupon the invention of new methods of inquiry. If quantitative findings cannot accrue, at least clarity and inclusiveness of observation and record may be attained." 30

### Procedural Precedent

There have been some nine studies which embark from a research philosophy and employ an approach approximate in many respects to that proposed here, namely the asking and answering of the question: where are we? where are we going? where can we go?

Three of these studies bear a broad relationship to the matters under consideration at the moment, because they reflect an awareness on the part of the investigators for the need for a relational analysis of the known-end-products of a body of research material or a method of research, or both, in their respective subject-matter fields of biology, education and sociology.

Pearl summarized and related the research practices extant in genetics as early as 1915 in his development of a sourcebook for research workers interested in applying systematic procedure to investigations in that field. 31 In 1922 the case history was analyzed as a method of research for use in education, 32 and again in 1939 the case method was studied in terms of its applicability to the realm of social research. 33

29. Gregg, Alan, The Furtherance of Medical Research, (New Haven, Connecticut, Yale University Press, 1941), 5. ("...the lack of rational doctrine and general interest in the problems of method has made medicine the scene of so much disunited and contradictory effort that it has caused it to be put down from its historical position as mother and nurse of the sciences.")

30. Holmes, H.W., Op.Cit., 174

31. Pearl, R., Loc. Cit.

32. Alltucker, H.M., Loc. Cit.

33. Whately, R.L., The Case-Study as a Method of Research, unpublished Doctor's Dissertation, New York University, 1931).

Original investigations which bear perhaps a more intimate relationship to both the philosophy and procedure being suggested here are to be found in the remaining six studies pertinent, for background purposes, to the present discussion. Davis summarized and analyzed reports of surveys of health and physical education and developed a checklist which he applied to the analysis of 117 city surveys of health education.<sup>34</sup> Odam compared methods of research reported by state teachers colleges, with the opinion of experts on the presumable value and direction of research in teachers colleges.<sup>35</sup>

Bixler constructed a series of checklists for educational research consisting of hundreds of items categorized broadly in the areas of selection and formulation of research problems, methods of research, techniques of research and presentation of research. His data was gathered from a survey of original research studies conducted in the subject-matter field of education.<sup>36</sup>

Atkinson analyzed techniques used in original research in five areas of the field of education, utilizing 540 published and unpublished doctoral dissertations and 150 published abridgements of research criticism drawn from the professional literature. This data was subjected to critical analysis by comparing the tendencies in original research with critical reviews of research procedures, in an effort to determine any relationship which might prevail. Research in educational measurement, educational administration, history of education and higher education represented the subject-matter phases which were examined to determine the method of research employed. The items examined in these fields yielded a 7-fold classification of general research method: experimental, analytical, distributive, predictive, historical and de-

34. Davis, E.C., Methods and Techniques Used in Surveying Health and Physical Education in City Schools (Contributions to Education, 515, Teachers College, Columbia University, New York, 1932)

35. Odam, G.A., Research in State Teachers Colleges and Normal Schools in the United States, unpublished Doctor's Dissertation, New York University 1932

36. Bixler, H.H., Check Lists of Educational Research (New York City, Bureau of Publications, Teachers College, Columbia University, Contributions to Education, 1928)

scriptive. Documentary-frequency counts were made on the data which represented a calendar span of 44 years of educational research.<sup>37</sup>

Taber was interested in surveying research in five fields of education for purposes of observing the tendencies in research techniques in original investigations, and to further determine whether or not the critiques of the procedures of educational research have been accompanied by a corresponding progress in practice. The data consisted of 416 doctoral dissertations and 200 published items of research subjected to a systematic analysis. This treatment of data was arranged to yield information on the problems and methods of research, the influence of critical discussions in the professional literature of research, and an evaluation of university-sponsored research in education. The five phases of education involved in this study included curriculum, learning, philosophy, teaching methods and guidance. The research examined spanned a 44-year period and was categorized as analytical, descriptive, distributive, experimental, historical, philosophical, predictive and relational.<sup>38</sup>

The adequacy of research techniques available for the solution of problems suggested in educational publications was considered by Englehardt, who surveyed much of the extant research in education which appeared in the form of published abridgements in professional journals. This study was conducted with the aim of noting the degree to which prevailing techniques for research could be usefully applied to problems typical of the subject-matter field of education.<sup>39</sup>

Gilkinson prepared and published an overview of research in general speech education for purposes of classifying and drawing together the end-results of investigations relating to effects of speech training, correlates of speech skill and speaker-audience relationships. Research method in this study was classified largely in terms of the criteria employed in various research items. These included preserved

37. Atkinson, A.M., Critical Analysis of Research Techniques in Educational Administration, Educational Measurements, History of Education and Higher Education, unpublished Doctor's Dissertation, The University of Pittsburgh, 1938. 400.

38. Taber, G.J., Loc. Cit.

39. Englehardt, M.D., "The Problems and Techniques of Educational Research," (unpublished Doctor's Dissertation, the University of Illinois, 1932).

samplings, comparison of initial and final ratings, initial and final personality test scores and counting of specific traits. Studies included extended across the subject-range of the field of speech with consideration given to theatre, radio, voice science, speech re-education, oral interpretation, experimental rhetoric and speech education. Items numbered 354, drawn from professional journals and largely postdating 1920. These studies have been interpreted in terms of the manner in which they appeared to add to, modify or contradict the body of generally-accepted information on the teaching of speech, for purposes of drawing together a recognizable body of research-based data about classroom methodologies as they effect the speech processes.<sup>40</sup>

### Conclusions

It is held that by some amplification of the procedures mentioned in this report, it is quite possible to probe the question of unifying educational research with the essentials of a scientific system and with the trends appearing to be uppermost in cognate socio-educational research methods, and that such a procedure will facilitate a pointing toward areas for future fruitful research. For an analysis as broad and inclusive as that being suggested here, the type of data used must be somewhat selective, because what is primarily needed at the moment are summaries of original research of the sort appearing in professional journals, and used as the basic data for at least one of the pioneering investigations mentioned above.<sup>41</sup> Examination of pertinent facts by this investigator and others has led to the broad conclusion that the main points of research design, procedure, statement of problem, description of data and conclusions have been summarized adequately by competent workers in the field.<sup>42</sup>

40. Gilkinson, H., "Outlines of Research in General Speech (Minneapolis, Minnesota, Burgess Company, 1943). 80.

41. Englehardt, M.D., Loc. Cit.

42. Another interesting precedent for the type of data being suggested here is cited by Spahr and Swenson who have pointed out that the matter of dealing with manuscripts and documents in research should not be an issue involving fine-line discrimination. A somewhat different point of view is presented by Methods and Status of Scientific Research (New York, Harper and Brothers, 1930), 6. Hull, C.L., Principles of Behavior (New York, Appleton-Century, 1943), Chapter 1.

One study completed at the University of Illinois and two Ph. D. dissertations completed at the University of Pittsburgh employ a research approach procedurally quite similar to that being implied here.<sup>43</sup> In one of these the data was drawn directly from the material appearing in professional journals relating to education.<sup>44</sup> In another, reporting on the status of unpublished doctoral-level research in education as compared with professional published research in education, Atkinson discovered that during 44 years of graduate work in the areas considered in his study, published research conformed to the minimum criteria which he set up. No single instance of unpublished research completely conformed to the minimum essentials which had been set up originally by the investigator as criteria to be used for examining research.<sup>45</sup>

Wolfle, Likert, Marquis and Sears in response to the report of the 1948 Policy and Planning Board of the American Psychologist, established the following tentative criteria for appraisal of extant psychological research in terms of its conceptual and experimental aspects: to what extent have first-hand observations which lead to testable hypotheses been made? what is the stage of theory development? are the problems which have been formulated--as well as the theories--stated in scientific terms? do the formulated concepts, theories and problems cover the area? how adequately can controlled observation of the variables involved in a given area be made? what is the level of mensuration in the field? how detailed is the knowledge of interaction among individuals? how well-coordinated is research in a given area?<sup>46</sup>

In each of these studies the investigator systematically classified and analyzed the data employed, each with a different outside criteria. In two instances the outside criteria took the form of critiques by experts on research method. In another the methods available were critically evaluated in terms of their adequacy of research design to meet specific problems posed for investigation in the field of education. In the fourth study mentioned above, the criteria were arbitrarily derived by a pooling of expert opinion, and have not yet been generally applied to the field of psychology.

In closing it should be noted that it is not within the province of this discussion to do more than mention in passing the broad designs for

43. Atkinson, Englehardt, Taber, Loc. Cit.

44. Englehardt, M.D., Loc. Cit.

45. Atkinson, A.M., Loc. Cit.

46. Wolfle, D., Likert, R., Marquis, D., Sears, R., "Standards for Appraising Psychological Research," The American Psychologist, 4:8 (August, 1949), 320 ff.

projects of the sort implied by this rationale. One fruitful line of investigation might be aimed at erecting a critical apparatus consisting of basic educational data, cognate research methods and the essentials of a scientific system, for application to the prevailing research practices and patterns. A second step would then be to proceed with the treatment of the research data of education in such fashion that the questions "where are we?" "where are we going?" and "where may we go?" can be answered rigorously and within a definite frame of reference. As has been noted above the American Psychological Association has already taken an initial step in this direction for purposes of restructuring psychological research.

Finally it seems reasonable to say that the time has come for some responsible agency to define educational research not alone in terms of function, nor alone in terms of scientific models, nor alone in terms of method of inquiry, but in terms of all such approaches, so that regardless of the broad critical apparatus eventually derived for such definition, future investigators may be provided with a full perspective of method, field and findings.

It is held here that this proposed redefining of research frameworks will do much to obviate criticisms like the following, which strikes directly at the nexus of the situation and states clearly the case for re-conceptualizing research structures:

"Too much attention has been given to the conducting of experiments. Einstein's equation, the atomic theory, the theory of evolution and the periodic table of elements outrank any five experiments one could name." 47

47. Ibid.

# RESEARCH GOES INTO ACTION

NORTON L. BEACH  
Teachers College  
Columbia University

TODAY AS YOU walk down the Main Streets of many American communities you are constantly aware of the shiny new buildings across the top of which appear two words: RESEARCH LABORATORY. American business has become more research conscious than ever before. Business leaders recognize that research is essential not only to improving business methods and practice but to survival itself. One leading industrialist from a midwestern state in chatting with the local superintendent of schools made this comment, "Mr. Superintendent, if we spent as little money on research in our organization as you do in your schools, we would be out of business tomorrow morning." Perhaps this school system should close its doors, but the problem is not quite so simple nor the picture so black.

There is growing evidence in the literature, in the policy deliberations of national organizations, and even more important in practice itself that a renaissance in educational research is under way.

It is particularly appropriate at this the most creative time in educational history when many ideas are being tried in scores of classrooms that an increasing number of individual teachers and groups of teachers are applying research methods to further test the quality of their practices.

## The Need for Research

The scientific movement in education has been under way for more than fifty years. During this period scientific studies have been the prime responsibility of the technician - the specialist skilled in using research techniques. Most often the work of these researchers has been on an individual basis in the university laboratory. To a large degree the results of this research during the last half century have failed to influence educational practice. There is some evidence to tell us the causes of this failure. As a rule the researcher was concerned only with the doing of a piece of research - he made little effort, after writing his story, to do the really hard job of helping to put the recommendations into practice. Furthermore, many of the research inquiries were directed toward finding solutions for problems that were not very vital to the potential consumer. The inquiries were usually conducted to establish broad generalizations.

Teachers, supervisors, and administrators have been busy in the last ten years of this fifty-year period trying out scores of practices. The solutions found in the research frequently had little to contribute to the testing of these newer practices. The practitioner, therefore applied his own powers of observation and study to determine the worth of the experimentation. This subjective approach has not always stood the tests of rigorous questioning. It is even possible that the present criticism of the public schools may be due in large measure to the fact that we can only defend what we do by saying "in our opinion" or "we have been doing this for X years." When pushed to the wall by citizens who want more objective evidence we are unable to give concrete facts in support of our position. Just the other day a meeting of superintendents was held where a significant piece of research was described. After the presentation was finished one superintendent after another urged the staff to complete the research program as quickly as possible and to find ways of making it available to superintendents, board members, and citizens as an antidote to growing community criticisms for which the professional staff had to answer.

Teachers, supervisors, and administrators are asking the question - what does research have to say about this? They do not always put the question in just this way at first but a brief discussion brings out the fact that they want tested answers. Men are looking to research to give them straight facts.

The need has grown so pronounced that local, state, regional, and national organizations have begun to attack the problem systematically.

The Metropolitan School Study Council more than ten years ago pioneered a movement that has resulted in the formation of some twenty-five school study councils organized on local and regional bases to pool their efforts and carry on a cooperative program in research. Within the last four years the Associated Public School Systems, a group of two hundred and five school systems organized on a nationwide basis, ~~has~~ begun to attack this problem.

The Association for Supervision and Curriculum Development has instituted a research program designed to assist local, state, and regional ASCD groups in stimulating, organizing, and carrying on local projects in cooperative curriculum research.

The Illinois Secondary School Curriculum Program is carrying forward a splendid statewide study which has achieved a number of first rate accomplishments in research at the secondary school level.

The Horace Mann-Lincoln Institute of School Experimentation has worked intensively with a small group of schools. This intensive program has made rapid progress to improve the quality of cooperative action research.

These are only a few of many illustrations where local, state and national leaders are giving recognition to the need for research and planning a design for meeting it.

### Change in Research Emphasis

Today there is a new look in educational research. Old patterns are being broadened and strengthened. Experimentation - a willingness to try a different approach - is more characteristic of the thinking and work of those doing research. Instead of borrowing without change research methods and techniques from the sciences, research workers in education are striking out on their own, sharpening their tools and building new skills.

What is happening is but a small beginning to be watched, encouraged and carefully evaluated. Let's look at what seems to be happening in the universities and public schools.

As mentioned earlier, until recently most of the research contributions have been made by individual researchers, professors, and graduate students, in the universities. This kind of research frequently called "traditional or fundamental research" by many writers is still an important part of our total research program. It will probably continue to be.

The composite picture of present day research programs in the universities might reveal the following classification of activities:

1. Research by individual professors
2. Research by groups of professors
3. Research by organized research bureaus or institutes
4. Research projects financed by foundations and government
5. Research developed cooperatively by the university and public schools, financed jointly with the public schools supplying "monies" made available through the local boards of education, and the university providing staff resources and facilities. The school study council movement is an important facet of this approach.
6. Research developed through the doctoral dissertations of individual students.
7. Research developed through doctoral dissertations completed as group projects.

It is interesting to note that there has been a major shift in the research usually initiated in the university. Individual professors, groups of professors, and the research bureaus and institutes are tackling specific researches where the roots are imbedded in the daily problems of schools. There is less of working on personal in-

terests and greater consideration is given to service research. Public schools are coming to the universities in larger numbers asking for research assistance. The cooperative survey, where the local school district furnishes assistance when needed, is one of the excellent outcomes of this major shift in emphasis.

One of the areas of research in the university that has grown tremendously since the war, partially because of the war stimulus, is the research project financed by a foundation or the government. A great many of these are directly concerned with the real and practical problems of the profession. The significance of the Cooperative Program in Educational Administration, supported by the Kellogg Foundation, for the improved training of school administrators is one of the best illustrations. This type of program has grown so rapidly and achieved such fine results that it has become a vital part of many college programs.

Within the last ten years the school study council movement has spread from one council to more than twenty-five. The school study council, which was organized to provide better opportunities for a group of schools to work cooperatively on their common problems with the university furnishing staff resources, has had remarkable success. These councils have made major research contributions in discovering solutions to the problems of their member schools. Many of these research contributions have become an important part of the educational literature and have been widely used by other school systems throughout the country.

For years the research work of individual graduate students in developing their dissertations has been an important part of the research picture. Recently a number of institutions have tried a new practice - that of having a group of doctoral students work together on a single dissertation. This type of group research has produced some excellent results according to reports from universities that have tried it.

In the past research was primarily a function of the university. Today and in the future many important research programs will be a shared responsibility of the university and the public school.

## ACTION RESEARCH

### General Characteristics

The last few years have witnessed a highly encouraging movement in bringing research closer to the problems of the school. As defined by Corey<sup>1</sup>

1. Corey, Stephen M. "Curriculum Development Through Action Research". *Educational Leadership*. December, 1949. Vol. VII, No. 3. P. 147

...this type of research that is conducted in local school situations and is designed to help the people working there know whether or not what they are doing is right is often called action research. The reason for this name is that investigations are made to determine the consequences of specific educational practices in actual school situations.

The theory of action research has spread rapidly and an increasing number of public school systems are attempting to inaugurate such programs as a way of achieving school improvement more rapidly. Basically, action research is directed toward the improvement of practices. This is quite in contrast with the major objective of fundamental research which is usually conducted to establish broad generalizations.

Caswell<sup>2</sup> has pointed out, however:

...that the action research approach must give careful attention, too, to the development of sound generalizations if the research results are to have meaning in other than local situations.

One of the chief criticisms of fundamental research has been that the person conducting the research has had little relationship to the specific project being studied. Action research offers a definite advantage in this respect because the people engaged in the research are the people carrying forward the school activities themselves. They are the ones who want to know the right answers to their practices.

More than twenty years ago John Dewey in The Sources of a Science of Education wrote convincingly of the need to engage teachers directly in research investigations:

It seems to me that the contributions that might come from class-room teachers are a comparatively neglected field; or, to change the metaphor, an ~~almost~~ unworked mine. It is unnecessary to point out ~~the~~ large extent which superintendents and principals have been drawn into the work of studying special problems and contributing material relative to them. It is hoped that the movement will not cease until all active classroom teachers, of whatever grade, are also drawn in.

There are undoubtedly obstacles in the way. It is often assumed, in effect if not in words, that class-room teachers have not themselves the training which will enable

<sup>2</sup>. Caswell, Hollis L. "Research in the Curriculum," Educational Leadership, April, 1950. Vol. VII, No. 7. P. 438

them to give effective intelligent cooperation. The objection proves too much, so much so that it is almost fatal to the idea of a workable scientific content in education. For these teachers are the ones in direct contact with pupils and hence the ones through whom the results of scientific findings finally reach students. They are the channels through which the consequences of educational theory come into the lives of those at school. I suspect that if these teachers are mainly channels of reception and transmission, the conclusions of science will be badly distorted and deflected before they get into the minds of pupils...

As far as schools are concerned, it is certain that the problems which require scientific treatment arise in actual relationships with students. Consequently, it is impossible to see how there can be an adequate flow of subject-matter to set and control the problems investigators deal with, unless there is active participation on the part of those directly engaged in teaching.<sup>3</sup>

Teachers, themselves, who have participated in action research projects are quick to agree with Dewey's comments. This kind of research has reached into only a few school systems and affected only a handful of teachers but its promise is great.

#### Advantages

Paul Mort has pointed out in several of his studies the great difficulty that schools encounter in getting research results into practice. The program of the Institute of Administrative Research, Teachers College, Columbia University, of which he is the executive officer, has had as one of its chief objectives the better utilization of research. Out of this experience one persistent conclusion emerges. School staffs that get together to study their problems in a scientific way are most ready to examine the experimental work of others. Action research is more likely to be used.

The ideas of many researchers will produce more creative lunches than the single investigator. Group thinking brings a variety of resources, background, and talent to bear on a specific problem.

Wherever two people come together to discuss mutual problems good ideas will be freely exchanged. Action research groups have helped to promote the philosophy of pooling and sharing.

3. Dewey, John. The Sources of a Science of Education. New York: Horace Liveright, 1929. pp. 46-48

There is a big job to be done in educational research. There are scores of studies to be made which will require the efforts of many people. The very nature of the action research approach makes it possible for a large number of persons to be involved. It should be pointed out that improvement in educational practice will continue to lag until a large number of individuals and groups are engaging in scientific investigation.

One of the striking results of conducting action research programs is that the members of the professional staff come more quickly to recognize that they need to get facts for consideration instead of generalizing from opinions.

### Limitations

There are many limitations in trying to work out action research programs. The solutions to most of these limitations are available in the results of one experience or another.

Teachers, themselves, say that the greatest difficulty they face is to find time to carry on the activities involved. Through the years we have continuously added duties to the teacher's load but seldom have we taken anything away. The school systems in the Metropolitan School Study Council have overcome this limitation to a large degree by providing substitutes for the teachers engaged in a study. This is not the only answer but administrators will need to develop some new inventions to provide time for teachers to participate in research. Perhaps boards of education and administrators need to recognize that top notch research work can only be produced when teachers are relieved of other responsibilities. It may be that action research groups should be organized during the summer and the staff members receive extra pay for these services.

For years the experts in research have pointed out all the difficulties and pitfalls in doing a good research job. Teachers are still gunshy. They are not sure about their participation. Reports of successful participation by teachers will encourage others.

Too often action research committees are established with little consideration given to the resource assistance that should be provided. Consultant help is necessary and where it is not available study groups are likely to accomplish less. They may even fail.

In any project that is just getting under way the quality of the product is often under par. This has been true of the results produced by some action research teams. The efficient use of research techniques requires time for mastery. Training programs might well be an answer to this problem. There is a need, too, for new techniques to be created where the old ones are not doing the job.

Research, whether it is done by an individual or a group, requires money. Some situations have been seriously limited because they have not faced up to this need.

The school climate is a critical factor in determining the success of an action research project. Administrative encouragement and assistance are vital. To date this has frequently been a serious limitation because administrators have not always recognized the importance of the role they must play.

### Values

The values of action research are far reaching in their effects on what happens in a school. Caswell<sup>4</sup> has explained it well when he says:

In the long run success in this area (research) is perhaps more vital than in any other in assuring the professional status of teaching.

As an increasing number of teachers and other professional workers contribute their share to produce good research more successful action in the classroom will result.

In the hearts and minds of many educators today there is the wish that colleges and public schools will find more ways of harnessing their potential strengths. Action research has opened the door to many avenues of shared responsibility and what is even more encouraging - the Main Streets of many American communities are alive with the hum of research laboratories in education at work.

4. Caswell. Op. Cit.

# SOME COMMENTS ON EDUCATIONAL RESEARCH AT MIDCENTURY

ARTHUR E. TRAXLER  
Educational Records Bureau  
New York, New York

## The Brief History of Research in the Social Sciences

IN ANY GENERAL appraisal of research in education we need, in the first place, to view educational research in historical perspective.

Since time immemorial human progress has been dependent largely upon the contributions of persons whose point of view and whose approach to everyday problems was essentially that of research--who observed accurately and objectively, generalized on the basis of observations, and applied the results of generalizations to specific situations. Thus, great inventions, such as the wheel, the bow and arrow, a written language, and other social institutions inherited from the period of prehistory, were without doubt products in part of rudimentary research.

The conscious and purposeful development and application of techniques of research is, however, essentially a modern phenomenon, even in the physical sciences. In the medieval university appeal to authority took precedence over direction observation as a means of learning. As late as the beginning of the seventeenth century, only a little over three hundred years ago, Galileo experimenting with the speed of falling bodies was a dangerous radical in a world of thought dominated by speculation and reference to the wisdom of the ancient world.

If the application of research to the understanding and control of the physical world is of comparatively recent origin, the use of research in the study of man and his development is indeed a modern process. There is very little which can properly be called research in any of the social sciences that can be traced back as far as the middle of the nineteenth century. Research in education is almost wholly encompassed by the twentieth century, and probably 99 per cent of the educational research studies now used for reference purposes were made within the last thirty-five years.

We are still in the exploratory and formative stage of a comparatively new science.

## The Nature of Educational Research--No Clear-Cut Boundaries

In truth, educational research is not an area separate and distinct from research in other sciences. It is a field without clearly defined

boundaries. Although it makes numerous contributions of its own, it borrows and applies techniques from a variety of other fields. In studies of physical growth it uses instruments and techniques of medicine; in investigating the legal status of the school it utilizes research procedures developed by the profession of law; in financial research it draws upon the practices of the accounting profession. In the long-term study of individuals it employs a case-study framework originally patterned after social case work. In countless studies of the intelligence, achievement, learning, and personal development of students, educational research uses not only the instruments and techniques but also the personnel of the field of psychology in such a close tie-up that numerous research workers may be classified as either educators or psychologists or both.

The overlapping of educational research with research in the other social sciences is inevitable because of the nature of the educative process itself. Essentially, education is the provision of conditions and the application of procedures to individual circumstances and needs in such a way that maximum development in relation to capacity will be fostered. Thus, everything that may bear upon the growth of an individual is potentially an instrumentality of education, and, by the same sign, a factor in educational research.

### Research and The Determination of Objectives

The primary need of man in the modern world--the one that far transcends any other--is the need for a set of ultimate values that will lend form and purpose to individual and group action. Not since the Holy Roman Empire of the Middle Ages has there been anything approaching agreement, even among the people of the Western World, on the great and controlling values of life, nor can there be, until there is developed a way of fitting the knowledge science has given us into a coherent philosophical pattern. As long as men remain confused and uncertain in regard to what life is all about there can be no worldwide stability.

It is debatable whether research may contribute to the determination of values or objectives. Many persons maintain that these are necessarily subjectively determined and that research must be confined to processes. However, in considering theoretically the whole function of research and in very comprehensive long-term planning, the possibility of devising research to help determine ultimate values and purposes should not be overlooked, although, at present, research is determined wholly or nearly so by values subjectively agreed upon.

### Broad Areas of Research

Most research in education can be classified into three very broad areas: (1) social foundations and framework of education, history of

education, and comparative education; (2) organization and administration of education, legal and physical aspects, and material matters of school plant and equipment; and (3) human learning, growth, and development and materials, methods, techniques, and personnel for evaluating and fostering that development.

By far, the largest number of research studies falls into the third area. This probably to be expected since it is this area which deals directly with persons, who are the focal point of all education.

If one were to try to identify any portion of one of these areas where the greatest progress in educational research has been made, he would probably settle finally upon the field of the basic skills. There is a tangibility and stability to objectives in this field which have tended to serve as reference points for research over long periods. It was to the field of the basic skills that Rice, Thorndike, Stone, Courtis, Judd, Freeman, and other pioneers in educational research first turned their attention, and in the ensuing years studies in this field have multiplied steadily. Improvement in basic skills is by common consent important at all school levels, and thus teachers have cooperated actively in research studies and have even carried on independent research of their own. Since investigation of the basic skills has been attractive to the amateur research worker as well as to the specialist, studies in this broad field probably have a greater unevenness of experimental design technique, and treatment of the data than those in other areas of educational research.

A second area of very extensive educational research is that concerned with the measurement of mental ability, achievement, interests, and personality. Particularly since the beginning of World War II, a large number of measurement specialists with competent mathematical backgrounds and extensive statistical training have been developed. Thus far, however, many of these highly trained measurement experts seem to a considerable degree to be preoccupied with the perfection of statistical techniques of measurement and have not yet brought their unusual training and ability to bear upon the study of fundamental educational problems. This is not to say that the perfection of statistical techniques is unimportant, but rather to stress the point that ~~it is~~ as a means to an end.

The study of mental growth and development is a related area of research where there has been noteworthy progress in comparison with the situation half a century ago. The analysis and characterization of individual pupils and school groups in terms of overall intelligence quotients, linguistic and quantitative aptitude, or standard scores on a profile of primary mental abilities utilizes a language much of which did not even exist at the turn of the century.

There are, however, many points of difference in the interpretation of data on mental growth, and recently the whole concept of measuring intelligence with non-culture-free instruments has been called into ques-

tion. If these instruments are named and employed according to what is probably their main use, the prediction of scholastic aptitude, there is probably less reason for concern over their culture-related character.

Notwithstanding the multiplication of studies in these and other areas, comparatively few definite unqualified statements of fact can be made on the basis of research. Most conclusions are guarded ones, stated in terms of probability. This is partly because it is the nature of the research personnel to be cautious; in fact, each research worker has an obligation to interpret his data cautiously and conservatively. But, in part, this tentativeness in conclusions is due to the nature of the material of educational research. When one is dealing with anything as complex as human personality--individually or en masse--seldom are definite, unqualified statements warranted.

### Some Characteristics and Needs of Educational Research

A large proportion of the research in education is done by graduate students working individually on comparatively narrow problems under the guidance of a faculty committee. These conditions are favorable to the training of a large number of young research workers, although it would seem that comparatively few continue to be very productive after they complete their period of graduate study. The circumstances of research by graduate students often lead to fragmentary studies, rather few of which are in themselves important for the findings they report.

Not infrequently, however, when a large number of these small individual studies concerning the same general problem are brought together, analyzed, and summarized, significant general trends emerge. Thus, summaries of research, of which the AERA Review of Educational Research and the Gray summaries in the field of reading are exemplary, are as valuable and necessary to progress in education as the original studies are.

Many questions in education can be answered with confidence only on the basis of long-term, longitudinal research. It has been usual to appraise the effect of varying environmental influences, or different instructional procedures, or remedial work of a particular design, or other variables on the basis of changes measured over short intervals. Generally, the question of whether the change is permanent, or more important, the question of whether anything is accomplished that would not in due course be accounted for by a normal growth process, is left unanswered.

The solution of numerous problems with regard to the development of the individual calls for either individual research workers who show an indefatigable and single-minded devotion to the long continued pursuit of a few specific problems with docile and patient subjects, or the conception and organization of research on a grand scale with financial support much more substantial than that now available.

Although many of the findings of educational research are tentative and uncertain, the potential values of research for school administrators and teachers are much greater than values now being realized. Many teachers have little or no knowledge of the research that has been done in their field and are unaware that there are practical findings that could be applied to their everyday work. There is a need for better publicity among school people with regard to educational research and for restating findings of research in nontechnical terms and in the language of the teacher. A step in this direction is being taken in the series of brief articles on research in different areas of education now being published by several members of the American Educational Research Association in the Nation's Schools.

But there is another aspect of the need to bring research more directly into the classroom. There has been a tendency among research workers themselves to regard research as something carried on under carefully controlled conditions by specially trained personnel, after which, it has been thought, the results could be transmitted to school personnel for reference and use in their work. In other words, research has often appeared to be something done for education rather than as a part of education.

As Corey and others have urged in recent years, there is an insistent need for the development throughout the schools of a conception of research, not as an independent set of techniques that can contribute to the teaching process, but as an integral part of that process. As research gradually becomes an aspect of everyday classroom practice with teachers themselves studying the effectiveness of their own procedures, it loses in erudition and refinement of techniques but it acquires a dynamic character difficult to achieve when confined to the work of the expert alone.

#### Important Areas of Future Research

Turning from the practical application of research to the everyday work of the school, one may inquire in what areas of educational research is especially needed in order to contribute to knowledge essential to the solution of some of the fundamental problems that beset our civilization. These are necessarily determined in the light of values. While agreement on a comprehensive set of basic premises cannot be assumed, the relationship of certain broad values or objectives to human welfare probably would not be seriously debated by anyone.

It seems self-evident that peace is to be preferred to war, health to sickness, economic and vocational security to insecurity, enlightenment to ignorance, and freedom of the individual to slavery, either economic or political, either physical or mental. It would seem that those processes by means of which peace, health, security, enlightenment, and individual freedom may be promoted are, next to the establishment of

a universal set of values, the most important areas of comprehensive research.

It is suggested that the following kinds of research are among the many kinds that might be chosen as appropriate to these areas:

1. Research designed to discover how the great mass of the people in all countries may be brought to a sufficiently high level of political intelligence so that they do not easily fall under the control of a dictatorship that uses power ruthlessly and cynically for its own ends. Since the school must furnish the main environment for this kind of training, educational research is of primary importance here.

2. Research directed toward the development of techniques for improving health knowledge and practices, including techniques of appraisal. This includes mental as well as physical health. Thus far, research has made only a meager beginning in this nebulous area.

3. Research in the long-term appraisal of techniques of individual guidance, particularly as applied to the vocations. In recent years, we have developed a considerable group of techniques for the guidance of individuals. Our procedures might begin to take on some of the aspects of a science if we knew how effective they were in the vocational adjustments of individuals after a period of years. Guidance of individuals toward vocational adjustment has great potentialities for the improvement of human welfare, but we shall have little real knowledge of the worth of our instruments and techniques until large-scale, long-term research is undertaken in this field.

4. Research directed toward the raising of the reading level and habits of the entire population. Notwithstanding the development of the radio and the rise of television, reading remains the primary vehicle of education. There cannot be an enlightened citizenship until everyone above the definitely subnormal level learns to read with understanding, learns to weigh printed statements and recognize propaganda, and learns to spend time reading and thinking about material concerned with important issues. There has been much research on reading in the United States, but a great deal of this research is fragmentary and inconclusive. It is doubtful if there is another field where sustained research liberally supported would pay off in such tangible dividends. Needless to say, it is desirable for research projects in this area to deal not only with the reading of English but also with that of other languages, including those of the Eastern World.

5. Research designed to devise and validate procedures for creating real understanding and appreciation of democracy among the youth of the democratic countries. We are notoriously behind the totalitarian states in techniques of developing understanding of and enthusiasm for our way of life. Somehow we have fallen, or are in imminent danger of falling, victim to a vast hoax. We have allowed democracy to be pushed into a position where it is considered conservative and reactionary

while communism is raised to catch the imagination of youth throughout the world as the ultimate in liberalism. Nothing could be further from the truth. Democracy, the idea of the maximum of individual freedom consistent with the orderly functioning of an organized society, is the most liberal, the most radical, idea yet conceived on this planet. Less than two hundred years ago, the radical and liberal character of democracy was understood much better than it is now. The idea of democracy has lost much of its drive. Somehow processes must be discovered for reviving its dynamic character. This is a goal to challenge the best that research has to offer.

### The Research Attitude and the Struggle for World Order

The processes of research are slow and painstaking, particularly in the social sciences--and the findings are oftentimes disappointingly inconclusive. Even a single research project may require years, and the accumulation of tested knowledge as study is added to study in any area, calls for scholarly endeavor through decades, generations, and even centuries.

In a social order stable over a long period, the slow working of research in the social sciences may be irksome to persons of impatient temperament, but in the normal course of events its contributions to human betterment probably cannot be seriously questioned.

In periods of great urgency--in times of worldwide turmoil and great national peril--there may be a considerable degree of doubt and confusion concerning the worth of research in education or any social science, even among the research workers themselves. The same long-term problems exist, and these may be closer to ultimate reality than any immediate problem, but beside crucial questions of group and individual survival in an armed world they are likely to seem at times academic even to their most ardent exponents.

The research worker in education looking at the troubled world to say is likely either to retreat into his own minute research studies or to experience considerable frustration as he considers the question, "What can I do? How can I use my training, experience, skills, and techniques in this pressing emergency?"

Among the possible answers that individual research workers might give to this question there is one that would seem to have universal applicability. Each person engaged in research can constantly strive to foster and to spread the research attitude.

In a world where subversion is a carefully nurtured art and science, where the sly and deft technique of the considered lie is probably the most potent modern weapon, not excluding the atomic bomb, its antidote, the attitude accompanying the relentless search for truth, acquires political importance. So, regardless of the findings of research, the

very process of research and the attitude accompanying that process are potentially a powerful force. The indoctrination of many individuals with that attitude could help materially in giving our people that intellectual and moral stamina essential for survival.

### Summary

In summary, I have tried to say that the history of educational research is comparatively so brief that we are yet in the formative stage of a relatively new science; that there are no clearcut boundaries to educational research but that it overlaps with several other fields, including medicine, law, accounting, social case work, and especially psychology; that it is conceivable that research may eventually be able to contribute to the determination of values or objectives but that at present research is formulated on the basis of values subjectively determined; that most educational research may be divided into three broad areas: social foundations, administrative, physical, and material matters, and human development and means of influencing it; that research on basic skills has been comparatively extensive and fruitful; that most conclusions in educational research are tentative, that thus far a large part of educational research is fragmentary and that there is a constant and pressing need for comprehensive, long-term research; that greater use of educational research among administrators and teachers needs to be fostered through summaries and nontechnical interpretations; that a conception of educational research as an integral part of the teaching process deserves much wider acceptance; that important areas for future research are the development of a higher level of political intelligence, the promotion of physical and mental health--knowledge and practice, the appraisal of techniques of individual guidance, the improvement of reading level and practices as a factor in an enlightened citizenship, and the development of more effective procedures for creating an understanding and appreciation of democracy; and, finally, that the development and extension among greater numbers of people of a research attitude--an attitude of search for truth--can be a powerful force making for stability and reason in a ~~fact~~ world.

# THE RESEARCH PROGRAM OF THE ILLINOIS CURRICULUM PROGRAM

ERIC H. JOHNSON, Illinois Curriculum Program  
CHARLES W. SANFORD, University of Illinois  
HAROLD C. HAND, University of Illinois

## Introduction

THE ILLINOIS Curriculum Program (ICP) is based on a thorough-going recognition of the fact that durable improvements in local public schools can be made only when the forces which support these improvements are local in character. The ICP is committed to the proposition that the public schools belong to the people. It actively endeavors not only to maintain but also to strengthen the time-honored tradition of local lay control of public education.

There is, thus, an underlying belief on the basis of which the ICP was designed and put into operation in 1947. This belief holds that enduring improvements will be made in public schools when, and only when, the people understand the problems which confront their schools and, with the direction furnished by specific basic information regarding the local situation, participate with specialists in working out sensible solutions to these problems. The central purpose of the ICP is to help local citizens, both lay and professional, to help themselves in the important business of improving their schools. The research program of the ICP is, therefore, focused upon giving assistance to local communities in developing understanding regarding school problems, making research procedures available whereby basic information regarding the schools may be discovered, and making special consultant help available to assist in working out sensible local solutions to curriculum problems.

It is virtually impossible to separate the research program of ICP from the total structure within which it operates. A brief description of the sponsorship, the purposes, and the activities of the Program is necessary to outline the framework within which the research program operates.

The ICP is officially sponsored by the Office of the State Superintendent of Public Instruction. The Program, therefore, operates under the auspices of the one agency that is most inclusively related to all of the schools of the state, regardless of level or support. The first purpose of the ICP has been to coordinate, under this sponsorship on a statewide level and on a local school level, all of the persons or groups who are, or who should be, interested in the school curriculum. On a state level, persons and groups have been effectively organized into a Steering Committee which includes representatives from

the State Superintendent's Office, institutions of higher learning, lay organizations and groups (including agriculture, labor, and business), and professional organizations of teachers and administrators. This group serves as an advisory committee to the State Superintendent on all matters relating to the ICP. At the local levels, the formation of similarly representative lay and professional committees has been encouraged.

The ICP has sponsored local studies which are basic to curriculum revision. These studies have been concerned with the holding power of the schools, the extent and character of extra-class participation, hidden tuition costs, guidance services, follow-up of graduates, the role of the schools in national security, and local area consensus development regarding the school's program.

Another purpose of the ICP has been to encourage local curriculum improving projects. All of the public and a substantial number of the private colleges and universities of the state and the Office of the Superintendent of Public Instruction supply to approved curriculum projects free consultative and other services to local schools conducting these projects.

Workshops on both a state and a regional basis have been conducted which have been attended by teachers, administrators, and lay citizens. These workshops have provided opportunities for these persons to meet together and, with the assistance of resource persons from higher institutions and from the Office of the Superintendent of Public Instruction, to work toward the solution of common problems.

Publications have been issued which have been concerned with manuals for conducting local basic studies, with the principal findings of the basic studies, and with general curriculum materials for use at the local level. These publications are distributed without charge to all schools in the state.

The ICP has also endeavored to establish improved relationships between the secondary schools and higher institutions. This work has been largely directed toward the modification of college entrance requirements in accordance with the kinds of competencies which are important to secondary school graduates for college success.

Attention here will be directed almost exclusively to these studies which have been designated as being basic to curriculum revision. As has been noted before, these research activities cannot be isolated from the total program structure, but an explanation of them will serve to illustrate the importance of their role in the development of the ICP.

### The Program

The fundamental assumption of the ICP that enduring improvements in the public schools should and will come about as the result of the

work of teachers, administrators, and lay citizens in local communities has been a very important factor in the design of those studies basic to curriculum revision which are popularly known as "basic studies." Another fundamental assumption, that the public schools exist to serve all of the children of all of the people, has likewise played an important role. A consideration of these two assumptions in combination points immediately to the most vital piece of information a local community needs in order to examine the program of the school. To initiate a curriculum study program at the local level, factual information needs to be available regarding the holding power of the school — the degree to which the school enrolls and retains to graduation all the youth of the community. If all of the youth of the community are not enrolled in the secondary school, specific information is needed regarding such questions as how many and what types of youngsters are failing to enroll, how many who do enroll remain in school to graduation, and what kinds of students typically drop out before graduation. The holding power study was developed to make available to local schools the techniques and the instruments necessary to secure this information. A how-to-do-it manual<sup>1</sup> for conducting the holding power study was published and made available at no charge to all secondary schools in Illinois so that any local community might utilize the material in assessing the local situation.

Two very important points regarding the holding power study need to be noted. The first is that emphasis is placed upon securing local facts. Many data are available regarding the holding power and the drop-out problems of secondary schools on a national, regional, and state basis. Such data, however, are not really meaningful to a local school and do not carry with them the impact necessary to stimulate local action. Only the facts concerning the local situation are likely to provide this impetus. The second point is that the data are useful only if they stimulate a local group to hypothesize regarding "why" a particular situation exists. Securing the facts concerning the holding power of the local school marks the beginning rather than the end of the local study. For example, if a local school finds that it is retaining to graduation only 48 per cent of the secondary school youth of the community, as was found in a certain high school in central Illinois in 1949, it can turn to other data yielded by the study for leads in formulating tentative hypotheses regarding why this situation prevails.

In the case of the high school noted, it was discovered that those students who dropped out had the following characteristics: (1) they

1. Allen, Charles M., How to Conduct the Holding Power Study. Circular Series A, No. 51, Illinois Secondary School Curriculum Program Bulletin No. 3, Office of the Superintendent of Public Instruction, Springfield, Illinois. May, 1949.

were dissatisfied with their school achievement; (2) they had more or less complex personal problems; and (3) they were from the lower economic levels.<sup>2</sup> The initial characteristic appeared to be associated with the curricular offerings of the school. Through curriculum-improving projects conducted with the assistance of specialists from higher institutions, attempts were made to reconstruct a number of the existing courses so that they were more nearly consistent with the needs and goals of students with lower achievement records.

The existence of the second characteristic indicated that the guidance services of the school needed to be examined. The ICP study concerned with guidance services was utilized at this point to examine the guidance program. This study had been developed and a how-to-do-it manual<sup>3</sup> distributed to enable local schools to study their own programs from the standpoint of how well they were meeting the needs and interests of the total student body. Data derived from this study were utilized to guide the revitalization of the local guidance program.

The third characteristic suggested that somewhere within the supposedly free public secondary school there might be costs of such a magnitude as to lead some students to drop out of school. A third basic study was used to test this hypothesis - the study of hidden tuition costs<sup>4</sup>. This study enabled the school to discover what it costs the student to take each course and to participate in the activities of the school. As a consequence of the study dealing with hidden tuition costs, steps were taken to alleviate excessive costs and to place virtually all of the activities on the free list. In a recent check-up (1952-1953 school year) it was found that the holding power of the school had increased to 82 per cent. Still another hypothesis as to why students were dropping out of school was suggested by the data of the holding power study; namely, that these youngsters were being excluded from the highly valued extra-class activities of the school. A basic study was developed which would enable any local school, through the application of procedures and materials made available

Sarf, Charles W., and others, The Story in Nineteen Schools, Circular Series A, No. 51, Illinois Secondary School Curriculum Program Bulletin No. 10, Office of the Superintendent of Public Instruction, Springfield, Illinois. September, 1950, p. 10

3. Lovelass, Harry D., How to Conduct the Study of the Guidance Services of the School, Circular Series A, No. 51, Illinois Secondary School Curriculum Program Bulletin No. 6, Office of the Superintendent of Public Instruction, Springfield, Illinois. May, 1949.
4. Hand, Harold C., How to Conduct the Hidden Tuition Costs Study, Circular Series A, No. 51, Illinois Secondary School Curriculum Program Bulletin No. 4, Office of the Superintendent of Public Instruction, Springfield, Illinois. May, 1949.

in a how-to-do-it manual,<sup>5</sup> to secure data bearing on this particular hypothesis. The information in this instance, as in all other basic studies, would be useful only as it served the local school in doing something about its own situation.

The follow-up study of the ICP utilized as its beginning point the data yielded by a poll of recent graduates, school patrons, teachers, and pupils in respect to their opinions as to what the school program as a whole should be doing to help meet the life needs of its students. This study too, was an outgrowth of one of the findings of the holding power study; namely, that nearly 80 per cent of the Illinois youth who dropped out of school appeared to have done so because they were not sufficiently persuaded of the value of their school subjects to give their best efforts to them.<sup>6</sup> Obviously, if these youth were to be kept in school beyond the legal school-leaving age, the subjects taught would have to be more visibly related to the life problems which they and their parents believed to be important. This at once raised the question as to what these life problems are. It is this question which the materials of the follow-up study were designed to help local patrons, pupils, and teachers answer for themselves.

Instruments were developed which, when administered to recent graduates, teachers, patrons, and students in the local community, would give an estimate of the prevailing climate of opinion with reference to the life problems of youth which should be met through the school program. As noted above the data derived from these instruments are not considered the end result of the study. Rather, they are utilized as starting points for local discussions of ways in which the curriculum of the school could prepare students to meet life problems more successfully.

For example, at a certain high school in northern Illinois the data derived from conducting the follow-up study were used as the basis for a series of meetings attended by school patrons. The implications of the data for the local program were discussed, gaps in the total learning experiences of students were ~~designed~~, hypotheses were formulated regarding why particular situations existed, and finally, long-range plans were developed which were designed to improve the local program.

5. Hand, Harold C., How to Conduct the Participation in Extra-Class Activities Study. Circular Series A, No. 51, Illinois Secondary School Curriculum Program Bulletin No. 5, Office of the Superintendent of Public Instruction, Springfield, Illinois. May, 1949.
6. Hand, Harold C., Principal Findings of the 1947-1948 Basic Studies of the Illinois Secondary School Curriculum Program. Circular Series A, No. 51, Illinois Secondary School Curriculum Program Bulletin No. 2, Office of the Superintendent of Public Instruction, Springfield, Illinois. May, 1949, p.14

The instruments for conducting this basic study are included in a how-to-do-it manual which is made available to all Illinois schools.<sup>7</sup>

As a direct outgrowth of the follow-up study, a series of action projects which have been designated as local area consensus studies, were developed. These studies are concerned, as is the follow-up study, with the purposes of the school. They are specific, however, with reference to the purposes of each one of the subject and service areas of the school.

Each one of these studies is designed to assist a representative panel of teachers, patrons, and students in any local community to do the following three things:

1. To come to an agreement as to what patrons, teachers, and students think the local school should be doing in the area under consideration;
2. To come to an agreement as to what patrons, teachers, and students think the local school is and is not doing in the area under consideration; and
3. To work out a locally-agreed-upon plan for doing a better job in those specific aspects of the school's program that appear to be in need of improvement.

The materials which are used to accomplish these purposes in any local school were developed by the ICP and are made available in quantity and at no cost to local schools. The development of materials for use in each one of the various subject and service areas (there will be approximately 22 of these when all of them are completed) constitutes a separate research project. It is necessary in each instance to develop a statement of the purposes which would, if fully implemented, lead to an outstanding program in the area under consideration. In most instances two people, one a specialist in subject matter and the other with a broad background in general education, are given the primary responsibility of preparing the statement of purposes. They are assisted by a panel composed of public school teachers, an administrator, one or more representatives from the Office of the Superintendent of Public Instruction, specialists from higher institutions, and lay representatives of lay groups. The instruments which are developed from the statement of purposes are used in communities as a starting point for discussions by laymen and teachers regarding the acceptability of these purposes for the local program. They are used, too, to locate those specific aspects of the local program which appear

7. Henderson, K.B., and Goerwitz, J.E., How to Conduct the Follow-Up Study. Circular Series A, No. 51, Illinois Secondary School Curriculum Program Bulletin No. 11, Office of the Superintendent of Public Instruction, Springfield, Illinois. August, 1950.

to be in need of improvement. Finally, planning aids are made available to local schools to assist them in the development of their plans for improving whatever it is that they decide needs to be improved.

The study which is concerned with the role of the school in the safeguarding of national security was developed to make materials available which might be helpful to schools in time of national emergency. This statewide study, involving over 2,500 lay and professional citizens of the state, resulted in a statement which identified the requisites for national security and made concrete recommendations concerning what the elementary and secondary schools should do to meet these requirements.<sup>8</sup> These recommendations have been widely considered by elementary and secondary school faculties, school boards, and parent-teacher groups. Through local discussion groups these recommendations have been examined for clues to the strengthening of the program of the local school.

### Summary

This description of the research program of the ICP illustrates the interrelatedness of the total undertaking. The basic studies were developed and are utilized in order to enable local communities to discover for themselves what problems they are confronting. Curriculum-development projects evolve out of these problems at the local level. Workshops and publications are used to assist local schools in carrying on these curriculum-development projects.

The research program of the ICP is best characterized as emphasizing action research. At all stages in the development and operation of the program the involvement of local citizens, both lay and professional, is stressed. The purposes of this program are to develop a wide understanding within the local community regarding the problems faced by the school and a broad basis of agreement regarding plans for improving the local program. The process does not stop with the discovery of basic local information but is initiated at that point. The real purpose of the research program is to secure the information that is needed to refine and accelerate the process of improving the schools.

The total program follows the most logical pattern of curriculum improvement. It involves the steps of determining who is being served by the schools, what is going on, what is needed, and how needed improvements may be brought about.

<sup>8.</sup> Sanford, Charles W., Hand, Harold C., and Spalding, Willard B., The Schools and National Security, Circular Series A., No. 51, Illinois Secondary School Curriculum Program Bulletin No. 16, Office of the Superintendent of Public Instruction, Springfield, Illinois, May, 1951. Copyright by McGraw-Hill Book Company.

# ACTION RESEARCH IN EDUCATION \*

STEPHEN M. COREY  
Teachers College,  
Columbia University

ACTION RESEARCH in education is research undertaken by practitioners in order that they may improve their practices. The people who actually teach children or supervise teachers or administer school systems attempt to solve their practical problems by using the methods of science. They accumulate evidence to define their problems more sharply. They draw upon all of the experience available to them as a source for action hypotheses that give promise of enabling them to ameliorate or eliminate the practical difficulties of their day by day work. They test out these promising procedures on the job, and again accumulate the best evidence they can of their effectiveness. They try to generalize as carefully as possible in order that their research will contribute to the solution of future problems or the elimination of future difficulties that they face in their teaching or supervision or administration. (1, 7, 17, 19, 20)

The practice of conducting research in order to improve decisions and practices that have to do with practical educational activities is not new. The term "action research", however, is relatively recent (2), and was first used by Collier (3), and Lewin and his students (10, 11.). Action research means much the same thing as "operational research," which was the name applied to a type of inquiry undertaken by various planning groups during the second World War (4). In education a great deal of the work done by the evaluation staff of the Eight Year Study is closely analogous to action research (16). Herrick (8) was referring to the same kind of inquiry when he cited the advantages of cooperative curriculum studies over curriculum surveys. The recent summary of Taba and associates (18) of the work of the intergroup education project of the American Council on Education describes numerous action research studies of intergroup relations.

One of the psychological values in action research is that people who must, by the very nature of their professional responsibilities, learn to improve their practices are the ones who engage in the research to learn what represents improvement. They, rather than

\*This article summarizes a longer statement by the same author entitled Action Research and the Improvement of Educational Practice, which will be published by the Teachers College Bureau of Publications, Teachers College, Columbia University, New York, N.Y.

someone else, try out new and seemingly more promising ways of teaching or supervising or administering, and they study the consequences.

There are two alternatives to action research as a method of improving educational practices. One is a procedure that most of us commonly resort to as we try to do our own jobs more adequately. Under the pressures of our work, we are disposed to change, if at all, on the basis of subjective impressions as to what our problems are. And with a minimum of emphasis on reality testing or the accumulation of relatively objective evidence, we form judgments about the consequences of our attempted improvements (6). As Lundberg has said, "... the scientific mode of thought is very recent in human affairs, ... is practiced by only a small percentage of our own generation, and... is uncongenial to a large number of otherwise admirable people." (12, p. 13)

The second alternative to action research is to ask the professional educational investigator to study our problems and to tell us what to do. There are two basic difficulties with this alternative. In the first place, the professional investigator can never study our problems in any strict sense. They are unique to the situations in which we are working, and we ourselves are part of the problem. A second limitation is that even when the recommendations make sense to us as we read them, we have great difficulty getting these recommendations into our behavior patterns (15, p. 53). We find it relatively easy to talk a better type of teaching or supervising or administering as a consequence of reading or hearing what others say that we should do. But there is a vast difference between this modification in our vocabularies and any substantial modification in the way we behave. An integral part of action research is our actual practice of the procedures that give a priori promise of enabling us to cope more effectively with our professional problems.

Probably the major difference between action research and the more traditional type of educational research has to do with the motivation of the investigators. In fundamental research, the basic motivation is to conduct an inquiry that will result in generalizations of widest applicability. In a real sense, the traditional educational research is motivated by his desire to arrive at the truth. If he is cautious and circumspect this truth is usually reported in relation to certain conceptions about probability. But to the degree that the investigation results in generalizations of the widest applicability, it is considered to be excellent research.

In considerable contrast to this kind of motivation, those who engage in action research, as has been said, are conducting their inquiries primarily because they wish to improve their own practices. Action research is conducted in the heat of combat, so to speak. It is conducted by teachers or supervisors or administrators in order

that they may have evidence as to whether or not they are accomplishing the things that they hope to accomplish.

The difference in investigational methodology between traditional research and action research are minor. Each type of investigator attempts to define the problem being studied with precision, to derive his hypotheses from as rich a background of information related to the problem as possible, to design an inquiry in such a way as to result in a genuine test of the hypothesis, to use facts or evidence throughout the research procedure rather than subjective impression, and to generalize cautiously and tentatively from the evidence collected. The situational circumstances under which the traditional investigator or action researcher carries out this process of inquiry, however, may differ appreciably. By various techniques the traditional investigator in education tries to exercise control over a situation so that many of the variables involved in real life teaching or supervising or administrating are ruled out by definition or by laboratory control. This practice results in a more definitive test of the stated hypotheses. But this precision is gained with a comparable loss in applicability of findings. People who conduct action research conduct their inquiries in the complicated psycho-sociological climate of on-going school activities. Because of the multiplicity of variables involved in these real situations, the action research inquiry is often lacking in precision, but in compensation it is realistic. The results have meaning for practice because they are a consequence of inquiry under life-like circumstances.

Whether the research be traditional or action oriented, its quality must be viewed in relative terms. It is possible to progress by small increments of improvements in procedure from the casual and impressionistic method that most of us use to cope with our practical difficulties to a methodology which is rather scrupulously scientific (6). In the degree that we are able to incorporate these increments of improvement in our action research methodology, we can have increasing confidence in the inference and generalizations that we draw from our research data.

These improvements in method of inquiry can characterize aspect of the total process of research. To expect to leap all from the method most of us employ as we try to do a better job method that incorporates the best scientific procedures is unrealistic. But to improve gradually requires little more than a commitment to the method of science as a way of coping with practical difficulties and a willingness to take the chances that are always involved in experimentation. This commitment to the method of science is more apt to result from the persuasion of events than from the persuasion of other men. In the last analysis, it must be the more dependable consequences of scientific problem solving that

argue most strongly for this method of coping with practical problems. These consequences are available for evaluation only when the scientific method is used. The best way to learn how to conduct action research, as well as the values it possesses, is to try it.

Action research in the field of education need not involve the co-operative activities of a number of people, but in most instances that is highly desirable. The reason is that a great many of the improvements that need to be made in teaching or supervising or administration cannot be brought about unless a number of people modify their perceptions and their practices and their points of view (14). Because this is the case, as many as possible of the people who will be affected by attempts to better a difficult situation should be involved in the action research on which the attempts are based. The cooperative efforts of all of these people will tend to result in better problem definition, a more realistic consideration of the numerous possible action hypotheses, easier translating of these hypotheses into action, and a more adequate understanding of the meaning of the evidence that is procured to test them (9).

Making action research cooperative in the above sense introduces all of the complications that are involved in group work. In few areas of human activity have precedents become more firmly entrenched. They have even been codified in manuals of rules of order for group meetings. For cooperative action research to be maximally effective, members of the working group must feel free to introduce modifications in their methods of procedure (13). Their ways of working must be appropriate to the job they are working on, and they must have numerous opportunities to try out and evaluate group procedures that give promise of increasing productivity.

If the quality of the methods now being employed by teachers and supervisors and administrators to improve their practices are to be bettered, there must be considerable change in the working environment and atmosphere of most school systems. Whether or not these changes will be effected, whether or not conditions will be established that are conducive to action research as a method of improvement, largely on things that the status leader is in the best position to accomplish (5). It is the school superintendents, principals, and supervisors who can do most to make it possible for school people to admit and discuss their own professional limitations, to hypothesize creatively and insightfully, to have the facilities and consultative help that are needed if these seemingly promising innovations are to be identified and put into practice, to get the best possible evidence on the consequences of these changes, and to derive from this evidence generalizations that are sound and provide helpful guides to future behavior.

There is considerable reason for believing that research methodology will not begin to have the influence on American education that it might have until thousands of teachers, administrators, supervisors, and school patrons make more frequent use of the methods of science as they cope with their own practical problems. In the last analysis, action research in education is no more than attempting to solve practical school problems by using research methods.

### BIBLIOGRAPHY

1. Banks, Tressa, Farley, Edgar S., Powers, Oscar, Vandermeer, Floyd, Waldorf, Robert, and Corey, Stephen M. "We Tested Some Beliefs About the Biographical Method." The School Review, 59:157-163, March 1951.
2. Chien, Isidor, Cook, Stuart W., and Harding, John. "The Field of Action Research." The American Psychologist, 3:43-50, February 1948.
3. Collier, John. "United States Indian Administration as a Laboratory of Ethnic Relations." Social Research, 12:265-303, May 1945
4. Committee on Operations Research. Operations Research with Special Reference to Non-military Operations. Washington, D.C., National Research Council, April 1951. 12 p.
5. Corey, Stephen M. "Conditions Conducive to Curricular Experimentation." Educational Administration and Supervision, 36: 209-216, April 1950
6. Corey, Stephen M. "Educational Research and the Solution of Practical Problems." Educational Leadership, 9:478-484, Dec. 1952
7. Foshay, A. Wellesley, and Hall, James A. "Experimenting in Moves Into the Classroom." Teachers College Record, 51:353-359, March 1950.
8. Herrick, Virgil E. "The Survey Versus the Cooperative Study." Educational Administration and Supervision, 34: 449-458, Dec. 1948.

9. Koopman, G. Robert, Miel, Alice, and Misner, Paul J. Democracy in School Administration. New York, D. Appleton-Century Company, Inc., cl943. 330p.
10. Lewin, Kurt. Resolving Social Conflicts. New York, Harper and Brothers, cl948. 230p.
11. Lippitt, Ronald. Training in Community Relations. New York Harper Brothers, 1949. 286 pp.
12. Lundberg, George A. Can Science Save Us? New York, Longmans, Green and Company, 1947. 122 p.
13. Mann, Floyd C. "Human Relations Skills in Social Research." Human Relations, Vol 4, No. 4, 341-354, 1951.
14. Miel, Alice. Changing the Curriculum: A Social Process. New York D. Appleton-Century Company, Inc., cl946. 242 p.
15. Mort, Paul R. and Cornell, Francis G. American Schools in Transition. New York, Bureau of Publications, Teachers College, Columbia University, 1941. 546 p.
16. Smith, Eugene R., Tyler, Ralph W. and the evaluation staff. Appraising and Recording Student Progress. New York, Harper and Brothers, 1942. 550p.
17. Smith, Mary Neel. "Action Research to Improve Teacher Planning Meetings." The School Review. 60:142-150, March 1952.
18. Taba, Hilda, Brady, Elizabeth Hall, and Robinson, John T. Inter-group Education in Public Schools. Washington, D.C., American Council on Education, cl952. 337 p.
19. ~~W~~en, Kenneth D. "Teachers as Researchers." Educational Leadership, 9:489-495, May 1952.
20. ~~W~~en, Kenneth D. Teacher Participation in Action Research Directed Toward Curriculum Change. Ed. D. Project Report. New York, Teachers College, Columbia University, 1950. 240 p. Typewritten.

# A PROPOSED TECHNIQUE FOR THE ANALYSIS OF DROP-OUTS AT A STATE COLLEGE

GRANVILLE B. JOHNSON, JR.  
Emory University, Georgia

IT IS of fundamental importance that a college fulfill the many needs of its students. Probably the most valid criterion of its success in this direction is the tendency for those students who begin a course of study to continue it through to its completion.

The present study was made to develop a method for ascertaining the degree of success which a particular state college had in attracting its freshmen students for a second year as sophomores, and to gain insight into ways of increasing this success. At the present time the problem is of particular importance because of decreased enrollment in many colleges, and it is felt that this proposed method may be of value to counsellors in other educational institutions where similar problems exist.

## Problem

The present investigation was carried out to develop a method significant to counsellors for ascertaining: (1) the seriousness of the problem of drop-outs at a particular educational institution, and (2) both subjectively and objectively the reasons for drop-outs after the freshman year.

## Subjects and materials

Subjects: The subjects were a representative sample of 1948 freshmen attending the state college in 1949 as sophomores and a like sample of freshmen who did not return after the freshman year.

## Materials

Tests: American Council on Education Psychological Exam, 1947 edition; Iowa Silent Reading Test, Form A. M.; and the California Test of Personality.

## Questionnaires

1. A questionnaire was sent to a representative sample of freshman students who did not return as sophomores to ascertain the reasons for drop-out.

2. A questionnaire was sent to a representative sample of colleges as similar as possible to the state college to find what percent of their freshman class did not return as sophomores.

#### Method of procedure

#### Extent of problems

1. The following percentages of freshmen were established.
    - a. The total of those who did not return to the state college the next year.
    - b. Men who did not return to the college the next year.
    - c. Women who did not return to the college the next year.
    - d. Anglos who did not return to college the next year.
    - e. Spanish who did not return to the college the next year.A representative sample of the total was made which constituted the experimental group.
  2. A questionnaire was devised and sent to 27 colleges of similar aim, size, background, within a radius as small as possible.
- The total drop-outs for the same period of the 17 schools returning questionnaires were compared with the drop-outs at the state college.

#### Subjective analysis

A questionnaire was sent to 50% of the freshman who did not return to the college the next year. Of the 125 questionnaires sent out, 44 were answered and returned. From this sample, the subjective analysis was made.

#### Objective analysis

Comparisons between the freshmen remaining and the freshmen who did not return were made on the following criteria:

1. American Council on Education Psychological Examination (quantitative and linguistic scores).
2. General Reading Test.
3. California Test of Personality (Self adjustment and social adjustment).
4. Grade point average.

#### Results

#### Extent of problem

Table I represents a breakdown of the college drop-outs from the

TABLE I  
PERCENTAGE OF FRESHMEN WHO DID NOT RETURN TO THE COL-  
LEGE THE NEXT YEAR

	Total	Men	Women	Anglos	Spanish
Total Freshmen	237	158	79	201	28
Freshmen drop-outs	162	109	53	135	22
% Freshmen drop-outs	68	69	67	67	79
C. R.		Insig.		1.65(.10)	



table that the difference between men and women is insignificant. Twelve percent (6% men and 15% women) of the freshman class of 237 students were bilingual Spanish students. 3.4% were neither Anglo nor Spanish. Since 67% of the freshmen Anglos did not return to the college the next year, and 79% of the freshman Spanish did not return (significance of the difference at .10 level) there is justification to conclude that counseling is not as effective for Spanish freshmen as for Anglo. This might be the result of the invalid reflective ability of the Anglo tests on individuals steeped primarily in the Spanish culture. The counsellor, too, may not be as skillful in coping with the personal and social problems of individuals from ethnic groups dissimilar to his own.

Of the 27 questionnaires sent to the various colleges, 17 were returned. The following table (Table II) depicts the significant data regarding them.

TABLE II  
ANALYSIS OF PERCENT OF DROP-OUTS IN 17 SELECTED COLLEGES\*

Range	M	OM	Mdn	
12-55	33.95	2.8	30.13	11.22 1.98

Analysis of these data reveals great diversity in the incidence of freshman drop-outs in the colleges which returned the questionnaire. If the sample of colleges may be considered typical (and there may be some doubt as to the validity of this presumption\*) the college under consideration is on the 99th percentile of freshman drop-outs for the year considered.

With reference to the foregoing data, it may be concluded that the freshman drop-out rate was abnormally high at the college and that this represented a real problem.

#### analysis

questionnaires sent to the freshman students who did not return to the college represented approximately 50% (random sample) of the group. Forty-four of these questionnaires (35%) were returned. There is a possibility that this sample was a selected one and the following results, shown in Table III, must be interpreted with this in mind.

\*This may not be a representative sample because the ten which did not answer may have had subjective reasons for not doing so.

TABLE III  
RESULTS OF QUESTIONNAIRE ANSWERED BY 44 FRESHMEN STUDENTS WHO DID NOT RETURN TO THE COLLEGE THE FOLLOWING YEAR

Reason for drop-out	Percent
1. Financial difficulty	25
2. Dissatisfaction with college	
a. Curriculum	20
b. Instruction	2
c. Administration	3
3. Personality adjustment difficulty	11
4. Marriage	9
5. Personal or family problems	9
6. Illness	7
7. Difficulty in meeting academic standards	2
8. Climate	5
9. Miscellaneous reasons	7



It may be noted from the table that the primary reasons elicited from the non-returning freshmen were financial and dissatisfaction with the curriculum at the college.

A high percent of the drop-out's reasons could be corrected by a counseling program which would assist the students in making academic, financial, personal and social adjustments, particularly during his first semester at the college.

### Objective analysis

An analysis of the four groups under consideration (men who stayed, men who left; women who stayed and women who left), reveals some rather interesting deviations in the six areas considered: (1) American Council on Education Psychological Examination (Quantitative score), (2) American Council on Education Psychological Examination (Linguistic score), (3) Iowa Silent Reading, (4) California Test of Personality (Social Adjustment), (5) California Test of Personality (Self adjustment), (6) Grade point average. The following conclusions were drawn with reference to these group differences:

#### Men who stayed and men who left:

The only statistically significant difference between men who stayed and men who left was to be found in the grade point average (.01). Approaching significance (.20) was social adjustment. These differences indicate that the men students who remained at the college had a superior grade point average and a tendency toward higher social integration than the men who did not return.

#### Women who stayed and women who left:

The most significant difference in this group was manifested on the Iowa Silent Reading (.02). Social adjustment and grade point average showed some tendency toward significance (.30). These differences would tend to indicate that those women who remained at the college were better readers, were better adjusted socially, and got better along their first year than those women who did not continue the first year.

#### Men who stayed and women who stayed:

There was no statistically significant difference between the men who stayed and the women who remained. The social adjustment of the women, however, most nearly approached significance over the men (.05). The women who remained also manifested greater self adjustment and reading ability than the men (.10).

#### Men who left and women who left:

Although the men tended toward superiority in quantitative ability

(.10) and the women manifested superior social and self adjustment (.10), there were no significant differences in the areas considered between men who left and women who left. Evidently they possessed similar aptitudes, abilities, and performance.

### Conclusion

It is possible to conclude from an analysis of the foregoing data that a high percent of the reasons for drop-outs could be corrected by a counseling program which would assist the student in making academic, financial, personal and social adjustments particularly during his first semester at the college.

Adjustment would be facilitated in probably more than just the academic area through the inclusion of a course in remedial reading.

Emphasis on the acquisition of insight into the specific problems and characteristics of minority groups would assist the counsellor in making more effective predictions when dealing with individuals from these groups.



# THE VALIDITY OF THE CHOICE OF FRIENDS-- METHOD OF MEASURING SOCIAL ADJUSTMENT

WILLIAM H. FOX, Indiana University  
DAVID SEGEL, Federal Security Agency

THE RESULTS of an over-all study of youth 14 and 17 years old (1)\* raised some doubt as to the value of the Moreno type (2) of social rating as an indicator of socialization or social adjustment. A hypothesis evolved from this study was that results of sociometric ratings of this type have little relationship to favorable social development in youth. This hypothesis is contrary to the present belief and practice of educators and psychologists.

The validity of the Moreno type of social rating was accepted largely on the basis of what might be called a case of "face" validity. That is, it was reasoned that the person or child who was chosen as a friend by a large number of persons or children was therefore socially acceptable and, consequently, socially adjusted. It is this assumption of relationship which is here questioned.

The problem therefore resolved itself into analyzing results of research studies bearing on the problem and adding such data as could be obtained through more intensive work of the over-all study mentioned above. It was found that the results could be presented in four categories as follows:

A. Studies the results of which show no relationship between the choice type of rating and other factors in youth. Besides the Moreno type sociometric rating, there were mental health scores, 1 averages of school marks, and several intellectual trait scores from the Differential Aptitude Tests that were available on the same groups in the over-all study referred to.<sup>2</sup> The correlation coefficients showing the strength of these relationships are given in Tables I and II.

\*Numbers refer to numbered references in the bibliography.

1. Published by the California Test Bureau, 5916 Hollywood ding dis-  
Angeles, California.

2. Published by the Psychological Corporation, 522 Fifth Avenue, New York, New York.

TABLE I  
 RELATIONSHIPS AMONG SOCIAL ACCEPTANCE RATINGS, MENTAL  
 HEALTH SCORES, AND HIGH SCHOOL MARKS  
 (72 Fourteen-year-old Boys)

Qualities correlated	Correlation Coefficient
Social acceptance and mental health	+.25*
Social acceptance and high school marks	+.28*
High School marks and mental health	+.55**

\*Significant at the 5 per cent level.

\*\*Significant at the 1 per cent level.

TABLE II  
 RELATION OF SOCIAL ACCEPTANCE RATINGS, MENTAL HEALTH  
 SCORES, AND HIGH SCHOOL MARKS WITH CERTAIN SUB-TEST  
 SCORES OF THE DIFFERENTIAL APTITUDE TEST  
 (72 Fourteen-year-old Boys)

Sub-areas of differential aptitude tests	Correlation with			High School Marks
	Social Acceptance	Mental Health		
Language usage--sentence	.04	.28**		.43**
Abstract reasoning	.07	.24*		.49**
Space relations	.08	.07		.32**
Numerical reasoning	.08	.28**		.51**
Mechanical reasoning	.14	.14		.16
Clerical speed and accuracy	.09	.36**		.36**
Reasoning	.05	.44**		.42**

\*Significant at the 5 per cent level.

\*\*Significant at the 1 per cent level.

The results show that mental health as here measured is related positively to school success and scores on certain intellectual traits which were available, but that the social ratings used have little or no relationship with these factors and with mental health as here measured. Goshorn (3) presents data from 533 high school students showing lack of relationship between this type of rating and Kuder Interest scores, while Greenblatt (4), using 32 girls and 33 boys in two seventh-grade classes, found no relation between a pupil's acceptance scores and mental health status and little or no relation between pupils' social acceptance scores and mental age.

B. Studies the results of which show positive relationship with other factors in youth. Several studies, such as those of Blanchard (5), Furfey (6), Seagoe (7) and Austin and Thompson (8), show that propinquity, i. e., accessibility for social intercourse, is fairly closely related to social acceptance scores. Austin and Thompson found that the most frequent reason given by children for making changes in friends was because of "lack of recent contact."

Another factor of possibly equal importance in the matter of choosing companions is the similarity of mental and other characteristics. Furfey (6) found that boys choose chums of the same size, age, intelligence, and maturity. Several studies, among them those of Hicks and Hayes (9), Jenkins (10), and Dimock (11), show that children and youth choose friends having the same traits as they themselves have.

C. Other types of sociometric ratings and their relationship to the Moreno type of social rating. Another type of sociometric rating is the type in which pupils are asked to judge the pupils as the basis of traits. Symonds and Sherman (12) and Havighurst and Taba (13) found a significant relationship between the results of the judgment type of rating and various types of personality judgment, including marks in school. Justman and Wrightstone (14) have studied the relationship among the Moreno type of social rating, a modified "guess who" test, and a modified form of the Ohio Social Acceptance Scale. The evidence and conclusions support the thesis that the modified "guess who" test, which asks for a judgment of the traits of pupils, measures somewhat different aspects of the social status of pupils than the other social rating scales which are variants of the choice, or Moreno social rating scale.

D. Other evidence. When pupils are asked to point out specific traits they liked in friends, they tend to select traits which may be classified as extrovert or socially aggressive. This is the conclusion of Kuhlen and Lee (15) and is shown indirectly from Cunningham's study (16) where the five top descriptive statements about best friends were as follows:

(a) "have fun with him or her," (b) "has other friends," (c) "easily liked," (d) "seems to come from a good home," and (e) "is a good sport."<sup>3</sup> If the qualities listed by Cunningham are studied, it may be seen that at least four of these qualities agree with Kuhlen and Lee's conclusion.

### Discussion and Conclusion

The lack of relationship between the choice rating type and other factors would of itself be not too disturbing, since it is possible that the results of this type of rating indicate a different kind of social adjustment from that yielded by mental health inventories, marks, and intellectual abilities. However, while these ratings are closely tied up with factors entirely extrinsic to social adjustment, such as social propinquity and the choice of friends having traits like themselves, it would seem that these social choice ratings are not actually related to any social or psychological factor within the individual child. The relationship of choice of friends to propinquity or likeness to oneself indicates nothing socially psychologically to either child who makes the ratings or the child who is rated.

The evidence concerning other types of ratings where judgment is asked for corroborates this finding in that it shows that, when a judgment or rating is made on the basis of a trait, pupils' ratings have some validity. This result gives a reason for the non-validity of the other choice type of rating.

The fact that pupils point out specific traits they liked in friends has no particular bearing on this problem since the choice of a friend and the selection of good traits in friends are procedures not of the same order.

Furthermore, the selection of traits tending towards extroversion does not necessarily mean that pupils select traits related to social adjustment. Extrovert traits can be undesirable also.

The writers conclude that the Moreno choice type of social rating has little or no relation to social adjustment.

ham, Ruth. Understanding Group Behavior of Boys and Girls. Publications, Teachers College, Columbia University, New York. 1933-185.

## BIBLIOGRAPHY

1. Ivins, W. H., Fox, W. H., and Segel, D., A Study of a Secondary School Program in Light of Characteristics and Needs of Youth. Bulletin of the School of Education, Vol 25, No. 6, Indiana University, Bloomington, 1949.
2. Moreno, J. L. Who Shall Survive? Nervous and Mental Diseases Publishing Company, Washington, D. C., 1943.
3. Goshorn, W. M. "A Study of the Relationships Between the Kuder Preference Record — Personal and Certain Sociometric Ratings," Abstract Ed. D. Thesis in Studies in Education, 1950. Thesis Abstract Series, No. 2, School of Education, Indiana University, Bloomington, 1951.
4. Greenblatt, E. L. "Relationship of Mental Health and Social Status," Journal of Educational Research (1950), 44:193-204.
5. Blanchard, B. E. "A Social Acceptance Study of Transported and Non-transported Pupils in a Rural Secondary School," Journal of Experimental Education (1947), 15:291-303.
6. Fursey, P. "Some Factors Influencing the Selection of Boys' Chums," Journal of Applied Psychology (1927), 11:47-51.
7. Seagoe, M. "Factors Influencing the Selection of Associates," Journal of Educational Research (1939), 27:32-40.
8. Austin, M. C. and Thompson, G. C. "Children's Friendships: A Study of the Bases on Which Children Select and Reject Their Best Friends," Journal of Educational Psychology (1948), 39: 101-116.
9. Hicks, J. A. and Hayes, M. "Study of the Characteristics of Junior High School Children," Child Development (1933) 5: 1-15. 
10. Jenkins, G. G. "Factors Involved in Children's Friendships," Journal of Educational Psychology (1931), 22:440-48.
11. Dimock, H. S. Rediscovering the Adolescent. Association Press, New York, 1937.
12. Symonds, P. M. and Sherman, M. Personality Services of a Junior

High School in the Measurement of Student Adjustment and Achievement. University of Michigan Press, Ann Arbor, 1949.

13. Havighurst, R. J. and Taba, H. Adolescent Character and Personality. John Wiley and Sons, Inc., New York, 1949.
14. Justman, J. and Wrightstone, J. W. "A Comparison of Three Methods of Measuring Pupil Status in the Classroom," Educational and Psychological Measurement (1951), 11:362-67.
15. Kuhlen, R. P. and Lee, B. J. "Personality Characteristics and Social Acceptability in Adolescence," Journal of Educational Psychology (1943). 34:321-40.
16. Cunningham, R. Group Behavior of Boys and Girls. Bureau of Publications, Teachers College, Columbia University, New York, 1951.

# A TECHNIQUE FOR EVALUATING STUDENT COMMENTS ABOUT GRADING PRACTICES

LAWRENCE P. BLUM and BEN A. SULLIVAN  
University of Wisconsin\*

IT IS NOT uncommon to hear from college students the complaint that certain teachers mark unfairly, and concomitantly it is difficult to determine if this is an emotion-charged rationalization or a statement of fact. Every campus has departments in which reputedly it is difficult for students to earn high grades and other departments in which grades are earned with less effort. It is desirable for college counselors to know which departments and courses are "low markers" and which are high. Knowing this, they can better interpret the comments of students. A simple and workable technique for assembling and presenting this information is presented in this report. The procedure is sufficiently general to be applicable to almost any college situation.

The specific object is to determine the comparative mean grade-point average for the institution as a whole, for the separate departments, and for the specific courses and sections of courses within each department. These levels are presented in outline form as follows:

## I. College as a Whole

### A. Departments as a Whole

(Example - the total program in English)

#### 1. Separate courses

(Example - English 1A)

a. Total of classes taught by one instructor

(Example - Mr. X might have three separate classes in English. For the present purpose, the three classes could be treated as one.)

Computations are made from the semester final grade reports. For the purposes of this study all courses are considered as carrying one credit. This is necessary because of the varying number of credits carried by students enrolled in the same course. An illustration of this is history which some carry for three credits and others, by extra work, carry four.

The steps in computing grade-point averages for the program in this study are as follows:

1. Grades of individual students including numerical weighting, B=2, C=1, D=0, E=-1/2, F=-1, are totaled for each section, course, department, and the institution as a whole.
2. The number of enrollments for sections, courses, and departments are totaled.
3. The number of enrolled students are divided into the total number

TABLE I

ILLUSTRATIVE PROCEDURE FOR COMPUTING GRADE-POINT AVERAGES FOR SECTIONS, COURSES, AND DEPARTMENTS

	Total Number of Enrollments	Total Number of Grade Points Earned	Mean Grade Point Average
<b>Department I</b>			
<b><u>Course 1A</u></b>			
Teacher X	75	150	2.00
Teacher Z	100	150	1.50
Teacher Y	25	50	2.00
<b>Total Course 1A</b>	<b>200</b>	<b>+</b>	<b>350</b>
			1.75
<b><u>Course 1B</u></b>			
Teacher X	50	50	1.00
Teacher Z	50	50	1.00
<b>Total Course 1B</b>	<b>100</b>	<b>+</b>	<b>100</b>
			1.00
<b>Total Department I</b>	<b>300</b>	<b>+</b>	<b>450</b>
<b>Total Department II</b>	<b>500</b>	<b>+</b>	<b>1000</b>
<b>Total Department III</b>	<b>400</b>	<b>+</b>	<b>400</b>
<b>Total Program as a Whole</b>	<b>1200</b>	<b>+</b>	<b>1850</b>
			1.54

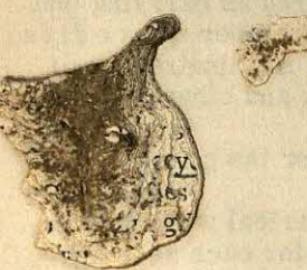


TABLE II  
ILLUSTRATIVE RANK ORDER OF DEPARTMENTS IN TERMS OF  
GRADE-POINT AVERAGES

Department Number	Grade-Point Average	+ or - from the School Average
1	2.09	+.70
2	1.68	+.29
3	1.53	+.14
4	1.30	-.09
5	.92	-.47

TABLE III  
ILLUSTRATIVE GRADE-POINT AVERAGES FOR COURSES WITHIN A  
DEPARTMENT

Department X

Department rank 6

Department mean-grade point average 1.96

Course	Rank	Mean grade-point average
64 A	1	2.4
53 A	2	2.3
52 B	3	2.0
52 A	4	1.8
53 B	5	1.6

ding dis-

ber of grade-points earned to yield a grade-point average or mean for each organizational level.

It can be noted on Table I that Teacher Z in Course 1A of Department 1 is grading well below the average for her department, but that her grading practices are in accord with those of the institution in general. Department II, however, seems to be grading higher than the institutional average. Table II depicts the manner in which this information can be assembled for use of counselors, department chairmen, and the institutional administration.

The course titles are included for the use of counselors and others who have need for information about all departments. For each department chairman the number of the specific department concerned is encircled. This enables each department chairman to determine the extent to which his department's grading practices are consistent with those for the institution as a whole.

Table III illustrates how this procedure would be useful for courses within departments and sections within courses.

A factor which limits the confidence one might have in the technique is that which is common to data involving averages. This involves the likelihood of distortion produced by grossly unusual grade averages at either end of the distribution.

A procedure such as is described can be useful to counselors in that it will assist them in easily distinguishing the realistic and factual content of counselee comment about marking practices from that which has no foundation in fact. It is useful to department chairmen and administrators because it permits comparisons of grading practices among departments, courses, and instructors.

## Research News and Communications

Lawrence P. Blum, University of Wisconsin Extension, Milwaukee; Jacob O. Bach, Southern Illinois University; Lester Sands, Santa Barbara College

Study composition of rural school boards in California. Matthew W. Ellerbroek recently completed a study at the Claremont Graduate School, Claremont Colleges, California, on the Socio-Economic Composition of Rural Boards of School Trustees. The study was conducted by means of a "schedule of information" which was sent to the school administrators of two hundred sixty-three school districts in southern California, with returns from 104 school districts.

Five conclusions were indicated by the study:

1. That the school trustees included in this study represent a narrow range of occupational, income, and educational classifications.
2. That the board members of large and small districts, according to average daily attendance, differ as to type of occupation, general level of income, and extent of formal education.
3. That there is an increasing representation of women on rural boards of school trustees.
4. That "new population" as defined in the schedule of information, has practically no representation on rural boards of school trustees.
5. That, according to the estimates submitted for the residents of the districts, a proportionate representation of the residents' occupational and income classifications is not in evidence on the boards of school trustees.

The study was under the direction of Dr. H. P. Allen at Claremont Colleges.

National Teacher Examinations to be held on February 13, 1954 at Princeton, N. J. The National Teacher Examinations, prepared and administered annually by Educational Testing Service, will be given at 200 testing centers throughout the United States on Saturday, February 13, 1954.

At the one-day testing session a candidate may take the Common Examinations, which include tests in Professional Information, General Culture, English Expression, and Non-verbal Reasoning, plus or minus of nine Optional Examinations designed to demonstrate his major subject matter to be taught. The college which a candidate is attending, or the school system in which he is seeking employment, will advise him whether he should take the National Teacher Examinations and which of the optional Examinations to select.

Application forms and a Bulletin of Information describing the procedure and containing sample test questions may be obtained from college officials, school superintendents, or directly from the National Teacher Examination, Educational Testing Service, P. O. Box 592, Princeton, New Jersey. Completed applications, accompanied by proper examination fees, will be accepted by the ETS office during November, and in January so long as they are received before January 15, 1954.

Study liberal arts needs and programs for the Wisconsin Lakeshore Area. An effort is being made to quantitatively and qualitatively define the needs for liberal arts education in the seven lakeshore counties within whose boundaries reside about one-half the population of the State of Wisconsin. The results of this and other studies will be used to guide future educational development in the area.

Three approaches to the problem were planned. The first consisted of compiling the liberal arts offerings of colleges and universities located in similar areas throughout the country. The second was a survey of the plans for higher education and the vocational plans of all graduating high school seniors in the area. The third was to survey the principal employers in the area to determine the need of employers for college graduates with liberal arts education. The first two approaches are completed and are summarized here. The third is now being completed.

After examining the liberal arts offerings of colleges or universities in similar areas, it was concluded that there existed very little agreement as to what should be included in a liberal arts program and very little uniformity as to what is included. An illustration of this is the observation that sixteen pre-professional programs were offered by the institutions studied with only three, pre-law, medicine, and dentistry, offered at all but one. Eighteen other areas were included in liberal arts programs, but without uniformity from college to college.

To determine the educational and vocational plans of high school seniors in these areas, questionnaires were completed and returned by 2,362 seniors. About 45% stated that they planned to attend college, 41% were not planning to attend, 14% desired to attend but were unable to, and the remainder were undecided. Of those planning to attend college 23% chose public-local institutions, 23% public non-local institutions, 24% private-local institutions, and 22% private non-local and out-of-state institutions. Nine per cent were undecided.

The career plans of the college-bound high school seniors indicated that 14% chose engineering, 14% business, 19% medical fields, 20% teaching, 21% miscellaneous, and the remainder were undecided. It was found that 75% of these students could meet their pre-professional requirements for one or more years at public local institutions.

Of the 463 students who wanted to, but were not planning to attend college, 52% gave financial reasons, 23% gave scholastic reasons, 10% stated military service, 16% gave miscellaneous reasons as family responsibilities, ill health, etc., and 16% gave no reason.

Occupational choices among this group were equally divided among teaching, medicine, and engineering with about 10% in each field. Business was chosen by 20% and miscellaneous fields by 12%. The remainder were undecided.

Students not desiring college attendance 47% chose some form of clerical work and 35% chose miscellaneous jobs spread over a wide variety of general titles. Eighteen percent were undecided.

Additional information concerning this study can be secured from Dr. Earl M. Grotke, Research Associate, Liberal Arts Project, Wisconsin State College, Milwaukee, Wisconsin.

# JOURNAL of EDUCATIONAL RESEARCH

---

Volume XLVII

February, 1954

Number 6

---

## SUMMARY OF READING INVESTIGATIONS: JULY 1, 1952 TO JUNE 30, 1953

WILLIAM S. GRAY  
University of Chicago

AS INDICATED by the unusually large number of references in the bibliography at the end of this article, interest in the scientific study of reading problems continues unabated. Ten of the references are devoted exclusively to summaries of published research relating to specific areas or problems. They have been included because of their value in guiding practice and in planning further research. Due to their character they should be read in detail by those interested in the topics discussed. Accordingly, they are referred to briefly in the statements that follow with no attempt to summarize their contents.

1. Readiness for beginning reading by Williams (118)
2. Differentiated instruction in reading, by Figurel (25)
3. The concept burden of instructional materials, by Serra (91)
4. How concepts and their verbal representations are acquired, by Serra (90)
5. Relationship between personality and the language arts, by Russell (86)
6. Relationship between personality and reading, by Robinson
7. The effectiveness of different types of graphic materials, by Malter (66)
8. Significant facts revealed by eye-movement studies, by (89)
9. Visual factors in reading disability, by Deady (16)
10. Is there a specific syndrome of visual factors in reading disability? by Shulman (95)

The chief findings of the new investigations reported during the year will be summarized under three headings: The Sociology of Reading; The Physiology and Psychology of Reading; and the Teaching of Reading.

Because of the large number of studies reported during the year, space will permit only very brief comments concerning each.

### The Sociology of Reading

Only a limited number of studies were reported during the year that were concerned with the social relationships and implications of reading. The fact was pointed out by Young (123) that what people read depends on what is available as well as their choice of the materials available. In order to find out trends in the nature of the materials published, he plotted the number of new titles, "with estimated corrections," published yearly from 1880 to 1949. Whereas there were wide fluctuations in the number of new titles yearly throughout the period, a straight line based upon corrected data revealed a highly significant trend. It showed an increase from less than 3000 new titles in 1880 to about 8600 new titles in 1949. When all the titles were distributed into sub-areas according to the International Classification System, it was found that fiction far outranked any other type. The likelihood is greater, therefore, that a reader will be able to find a book of fiction that he would like to read than a book in any other field. The areas most widely represented by new titles, in addition to fiction, were religion, juveniles, literature, and philosophy, sociology, poetry and drama. An analysis of all the facts secured led the investigator to conclude that "the books of a year (always) reflect its important events and questions."

Believing that comic strips "are complex stimulus patterns in which a variety of positive and negative values are expressed" Spiegelman, Terwilliger, and Fearing (102) made an analysis of "goals and means to goals of comic strip characters" over a three-week period in 52 nationally syndicated comic strips. The findings showed first that the world represented by comic strips is predominantly a man's world. The male characters belong largely to the lower middle and lower socio-economic class, more female characters are in the upper-middle and middle classes. Male characters. They showed, second, that male characters in the upper class are generally humanitarian in both goals and means, that those in the middle class "seek self and power goals through aggressive means," and that the lower class "seeks both altruistic and personal goals depending on fate and charm as means." The

findings also that female characters of the upper class are generally concerned "with being loved and lovable," the tendency of the middle class is "to achieve career recognition by hard work," and the tendency of the lower class is "to serve others and accept their fate." To what extent these goals and means are representative of the respective classes in real life was not considered.

Malter (66) made an analysis of the content of 185 comic magazines. The findings showed that about 25 percent of the space in them was de-

voted to covers, advertisements, and special features. The remaining space was devoted to the following types of stories: 11.1 percent, western; adventure, 10.7 percent; annual, 10.3 percent; love, 10.2 percent; detective, 9.3 percent; "superman", 6.9 percent; adult antics, 6.6 percent; adolescent antics, 5.2 percent; jungle stories, 2.6 percent; children's antics, 2.3 percent; and sport stories, 0.5 percent. As a result of a detailed study of the nature of the various classes of content identified, Malter concluded that "it seems unreasonable to blanket all comic magazines under the heading 'unacceptable'." As constructive measures, he recommended that effort be made to eliminate unacceptable comic magazines by teaching children to be more selective and by making other acceptable books more available.

Tannenbaum (108) investigated the role of the headline in everyday newspaper reading behavior. He found through experiments with college students that they "exert a significant influence" upon the opinions of readers. The headline "sets the stage, as it were, for the manner in which the story is read." Thus the headline writer may play a highly significant role in determining how the news is interpreted.

Maclean and Hazard (58) studied the appeal made by 51 pictures to 152 women. Interest ratings were first secured. The content of the pictures were then analyzed as were the comments about pictures made by those who ranked them "very high" or "very low." Six major types of appeals were identified: "idolatry"; "social problems"; "pictur-esque"; "war"; "blood and violence"; "spectator sports." It was believed that the findings were sufficiently significant to experiment further with the use of these appeals as aids in directing the attention of women to social issues and other items of large public importance.

#### The Physiology and Psychology of Reading

A wide diversity of problems relating to the physiology and psychology of reading were studied during the year.

Is reading a single complex ability or does it differentiate in series of complex activities? Studies by Shores and Sause (94) led to the conclusion that "reading ability differentiates beyond the period of a year from the third to the sixth grade into somewhat specific abilities to read different kinds of material for different purposes." Conclusions were based on intercorrelations among scores made by 214 fourth-, fifth- and sixth-grade children on five tests—science reading, California Mental Maturity Scale (with language and non-language sections), reading, and arithmetic. The investigators believed that with measuring instruments of greater discrimination and precision the above findings would be substantiated and, in addition, the extent and nature of the differentiation and the amount and character of the remaining common factors would be described.

Relation of reading ability to other factors. The foregoing study showed that reading ability and mental age are positively correlated. Clark (9) gave further emphasis to this fact by charting the scores on the reading section of the California Achievement Test made by pupils having different I. Q.'s at each grade level from one to fourteen. The results supplied convincing evidence of a positive correlation. In a study of Welch-speaking children, Jones (49) found that there was a steady increase in reading age with increase in I. Q.'s on a non-verbal test.

The fact has been emphasized repeatedly during recent years that reading ability is influenced by language mastery. Jones (48) conducted an experiment in speech training at the third-grade level which led to the conclusion that "speech training does affect positively the silent-reading achievement of pupils" and that "speech and reading are closely related aspects of language." In a study including both monoglot English-speaking and bilingual Welch children, Jones (50) found a "highly significant difference" in favor of the former group on both the verbal intelligence and the silent-reading tests used. The explanation offered was that the bilingual group had not yet acquired the ability to think in English with the facility and accuracy of the monoglots, due to "their linguistic environment both in and out of school."

In a study of the relationship between auditory characteristics and specific silent-reading abilities, Reynolds (80) failed to secure evidence that auditory ability, as measured by a series of carefully-selected tests, was related to general reading ability. He found on the other hand that scores on tests of auditory memory span, word discrimination ability, and pitch discriminatory ability may be used to advantage in predicting word recognition ability and capacity to learn the sound values for common word elements. Somewhat different results were secured by Poling (76) who compared the scores of good and poor readers on various auditory tests. She found significant differences in their auditory memory span but not in auditory acuity or discrimination.

<sup>the</sup> relationship between visual characteristics and reading efficiency has been studied widely in the past. During the current year, Edson, Cook (19) reported the results of a relevant study involving fifth-grade pupils. It aimed to determine the "relationship between variations in ten measures of silent-reading skills and thirteen visual characteristics." The findings indicated that "within these groups and for the tests employed no evidence was found to lend support to the opinion that achievement in reading is limited by vision." The fact was emphasized, however, that every effort should be made "to insure visual comfort through adequate attention to visual defects by qualified persons." Through the use of a checklist of visual difficulties, Knox (52) found that the number of different symptoms of visual defects exhibited by pupils was not a good criterion for referral to

a refractionist. Not infrequently other factors, such as emotional maladjustment, are more significant in causing reading difficulty than the visual defects noted.

In a detailed study of "visual efficiency and learning to read" Robinson and Huelsman (84) gave to pupils in grades I, IV, and VII a battery of visual tests and of reading tests. Intercorrelations among individual scores "revealed very slight and inconsistent relationships." However, the results of a group factor analysis showed a "relationship between composite reading achievement" and the results of tests of near acuity as well as of depth tests. Additional studies were suggested for securing a clearer understanding of the relationships between visual performance and reading progress and to identify valid and reliable tests for school use.

That many aspects of personality exert an influence on reading achievement was emphasized by Solomon (100). The Rorschach Test was given to 45 children shortly after beginning the first grade and again in the third grade. An analysis of these records, together with those relating to progress in reading, indicated that successful learners tend to emphasize the abstract or theoretical in their approach to problems while the less successful give more attention to unimportant details. The former group exhibit greater accuracy of perception of their environment and "show greater feelings of anxiety and inadequacy." Successful boys tend to be more introspective than unsuccessful ones and successful girls are as a rule "a well-adjusted group emotionally and intellectually." Lack of success in learning to read was accompanied by such changes in personality patterns as "a marked increase in the quantity of affective energy and capacity for environmental contact" and an increase in "immature, impulsive, emotional reactions."

In a study involving 24 students in educational psychology, Roesch (85) secured introspective evidence to the effect that interest in a task, effort, and "the will to learn" were vital factors in overcoming difficulties and in making progress in such tasks as vocabulary mastery, increasing speed of reading and learning to read Latin scripture.

Relation to left-hand orientation in reading to other activities  
In order to determine if left-right orientation in reading influenced a tendency in drawing, Jensen (45) studied pertinent data secured from cultural groups — two American populations, one in Norway and in Egypt. His findings showed that "regardless of the direction of the established reading habit or of the amount of practice in reading, the tendency to draw profile facing toward the left was present in all cultural groups." In a similar study involving Japanese children, who read from top to bottom and from right to left, Jensen (46) found that "the proportion of profiles facing to the left was significantly greater than chance expectancy."

Word perception. Several investigators were concerned with prob-

lems relating to word perception. The chief findings in some of them can be summarized readily by stating the hypothesis under investigation in each case and the conclusion reached. Forgays (27) secured data to test the hypothesis that differential recognition of words to the right and left of central fixation is largely the result of training. His data showed that "superiority of recognition of words presented to the right of fixation over the left is significantly related to educational grade level." These findings were interpreted as supporting the assumption. Haigh and Fiske (36) were concerned with the assumption that personal values which the reader attaches in words greatly influence perception. Use was made of stimulus words belonging to six value categories. When presented tachistoscopically, they found that shorter recognition times were associated with words of high value and longer recognition times with words of low value.

Several investigators sought evidence which would support or deny the assumption that the perception of words is related to some extent to frequency of word usage. McGinnies, Comer, and Lacey (60) found that despite differences in visual acuity, regression equations for individuals indicated that duration thresholds for neutrally toned words are a "linear, decreasing function of word frequency and a linear, increasing function of word length."

Using non-sense syllables as stimuli, Solomon and Postman (99) secured evidence which supported the assumption that tachistoscopic recognition is facilitated by prior usage of the words. Cofer and Shevits (10) found a positive relationship between the frequency of occurrence of words in usage and the number of associations given to stimulus words. They pointed out that whereas their findings "in no way demonstrate that the recognition-time personal-value relationship is predicted upon the personal familiarity or frequency of the stimulus words, they are consistent with this interpretation." Lepley and Kobrick (57) secured evidence that the number of synonyms attached to a concept by an individual is related with the frequency with which he used the concept. All these findings emphasize directly or indirectly the importance of "personal values" in word recognition.

Relationship among vocabulary, structural analysis and reading. In a series of scores made by college students on a standard reading test on tests constructed to measure vocabulary, rate, comprehension and structural analysis, Hunt (43) reached the following conclusions: vocabulary, structural analysis and vocabulary are "moderately" interrelated; reading and vocabulary are more closely related than is structural analysis to either of them; of the four measures of structural analysis, "context is most closely related to vocabulary and reading"; the more intelligent subjects make more effective use of structural analysis than the less intelligent; however, structural analysis is more than intellectual capacity "since it tends to be related to vocabulary and reading after the effect of intelligence is controlled."

A final conclusion was that effort to train pupils in methods of word attack should be encouraged.

Nature of the meanings attached to words. In order to secure a clearer understanding of how meanings are acquired, Werner and Kaplan (116) developed a "word context" test. It was given to 125 children, ranging in I. Q.'s from 101 to 111, who were distributed equally among five age groups from 8.6 to 13.5. A detailed quantitative and qualitative analysis of the test records led to the following conclusions among many others: growth in word meaning increases steadily from grade to grade; "complete-incorrect solutions" decrease rapidly from the first to the second group and remain more or less constant thereafter; variability in the interpretation of test situations decreases with age showing the influence of socialization on language behavior; at immature levels, a word does not possess stability, but may have "a wide and often diffuse contextual connotation", may be "fused with other concepts" and "its meaning may be readily altered"; immaturity is also characterized by failure to distinguish a word grammatically from "a larger semantic unit" and by lack of "definiteness of syntactic structure"; the relation between symbol and referent becomes increasingly indirect with age. A high degree of abstract symbolism indicates that "the child has learned to differentiate between areas where a relatively concrete symbolism is quite in order, and areas where a hypothetical attitude is required for adequate performance."

Further study of the meanings which children attach to words was made by Kruglov (54) who gave multiple choice vocabulary tests to pupils at the third-, fifth-, and seventh-and eighth-grade levels. Significant differences were revealed. Younger children tend to select "more use and description and more repetition-illustration-inferior explanation type responses" than older children. Furthermore, the younger children tend "to choose more concrete definitions and the older groups more abstract definitions" and older children "tend to choose more synonyms" than younger children.

Recognition and retention of once read material. Two different investigations were reported in this field. Sister Columba Mullaly measured the retention of information learned from one reading of a prose passage by 1,278 pupils in grades five through eight. A test, and retest at varying intervals were given. The results of initial tests showed that in general about 35 per cent of the items were learned from one reading. No sex difference was revealed but there was a definite gradation in the number of items learned from bright to dull pupils. The percentages of retention evidenced at various intervals following the first reading supports the prevailing belief that "forgetting takes place rapidly at first and more gradually as time elapses." Reminiscence "tends to compensate for forgetting", the poorer learners resorting to it much oftener than the brighter ones.

Sharpe (92) conducted an experiment similar to that of Sister Columba. It was based on the same test material and the same items of information, the chief difference being that it was "a study of the conditions of recall." A total of 2,492 pupils, varying in age from nine to eighteen were involved. An analysis of the data secured showed (1) that total achievement rose during the first-day interval with a gradual decline toward later intervals; (2) that "forgetting takes place as a gradual process, and is not an abrupt falling away as the practice ends"; and (3) reminiscence increased on the first day, "with an irregular advance through later intervals."

Speed of reading and skimming. Using college students as subjects, Perry (74) attempted to find out the comparative ease and accuracy of reading Arabic and Roman numerals. The subjects were allowed one minute for each test and were told to read as fast and accurately as they could. The speed of reading the Arabic numerals, 1 to 9, was 50.1 per cent greater than reading the corresponding Roman numerals, 137.5 per cent greater for numerals from 10 to 49, and 349.4 per cent greater for numerals from 50 to 99. With respect to accuracy the corresponding per cents, in favor of Arabic numerals were 75, 96.4, and 97.1. Mitchell (71) experimented with a battery of tests to find out if tests could be identified that would help to determine if improvement in speed of reading might be expected through training on the part of a slow reader. The results did not justify conclusions but gave encouragement to further study of the problem.

Grayum (35) carried on a study of skimming involving six different age groups varying from fourth-grade pupils to adults. The results of tests, controlled observations, and eye-movement records were used in determining the nature of the skimming process. The findings revealed wide differences among subjects at each grade level in ability to skim. The chief processes identified were: "skipping in various degrees"; "marked changes in the regular reading rate"; "pausing"; "regressing"; "looking back"; "looking ahead". In addition, each syllable who looked high in skimming ability made use of detailed processes "of their own devise."

Interests in leisure-time activities, reading, television, and radio. In order to find out "what children like to do but do not do and the accompanying reasons," Sullenger, Parke and Wallin (105) made a survey of the leisure-time activities of children in grades four to eight in an industrial area of Omaha. From the point of view of this summary the most significant finding was that reading ranked third in order of preference among the twenty most popular activities. Listening to the radio and attending movies ranked first and second respectively. Among the so-called "quiet-activities" children "spend a large part of their leisure-time in reading." Approximately 82 per cent reported that they enjoyed reading. The findings of Sheldon and Cutts (93) were less

favorable to reading. They reported that "almost half of the above-average and the superior readers have reading as an out-of-school interest or hobby" and "about one-fourth of the average and only one-tenth of the below-average readers seem to be interested in reading at home." The differences in the findings of the two studies just reviewed are typical of variations usually found among communities. A study by Eppel and Eppel (22) of the interests of young workers at a county college in England showed that reading ranked eighth from the top among boys in a list of 19 leisure-time activities and seventh from the top among girls in a list of 24 leisure-time activities.

Evidence of wide appeal which radio and television makes to high-school students was secured by Maz (69) who analyzed 721 questionnaire returns from high-school students and their parents in Pittsburgh. The average time spent per day in televiewing by boys who lived in television homes was 3.2 hours and in listening to the radio was 2.1 hours. However, if there was no television set in the home 3.8 hours per day was spent in listening to the radio. Life and Readers Digest were the two magazines most preferred by both students and parents. A significant conclusion was that "comparatively few read serious material, a few more will hear it by way of radio, but the largest potential audience is there waiting to be served by educational television."

Replies secured by Witty (120) from 41 individuals who had formerly appeared three or more times on "quiz kid" programs revealed the fact that they read widely. The chief value which the respondents attached to reading was that it aided them "in formulating a philosophy of life." Emphasis was also given to its value in securing a better understanding of self, nature, society and the demands of certain occupations. The high rank which bright children attach to reading was further evidenced by the findings of Cutts and Mosely (15) which showed that reading was exceeded in preference in the case of only one subject, by bright pupils in grades II to XII. Brown (6) studied one area of the reading interests of college students analyzing their readings in "current science." The findings showed that they read chiefly in non-technical areas of the biological sciences and concerning problems that have "a direct relationship to themselves or to mankind."

Readability. Wide interest continues in the factors which influence readability, in the validity of existing readability formulas, and in their application to reading material at all levels. Klare (51) reported an evaluative study of the Gray-Leary Flesch and Dale-Chall formulas for measuring the readability of written materials. The findings were based on an analysis of 52 books and three reading tests. The investigator concluded that these formulas measure about the same aspects of adult readability, that the Dale-Chall and Flesch formulas are most similar, that the Dale-Chall is probably superior to the other two, and

that grade ratings which are given to children's books by such formulas as the Washburne-Morphett and the Lewerenz have questionable meaning when applied to adult materials.

Two studies reported the results of inquiries concerning the reliability of the original and the simplified Flesch Reading-Ease Formula. England, Thomas and Patterson (20) presented evidence which supported the contention that "both the original and the simplified Flesch reading ease formula are highly reliable." Dunnette and Maloney (18) reported the results of a study to determine if the new Flesch formula "is operationally a simplified version of the old or whether it is simplified in name only." Their conclusion was that "the revised formula is superior with respect both to time taken and accuracy with which it is applied."

Swanson and Fox (106) were concerned with the value of the use of readability formulas at the adult level. Their findings led to the conclusion that they can be used "to predict differences in comprehension between two versions of the same material." However, the ratings given by the formulas used did not prove to be valuable aids "in predicting differences in readership" or retention of material. Certain factors limiting the conclusiveness of the findings were pointed out. Tarleton (109) applied the Flesch formula to 40 science stories presented through mass media to determine if simple material is as accurate as more difficult material. Three of the four inaccurate stories were rated as difficult. Further study showed also that the length of an article has little influence on its accuracy. The investigator emphasized the fact that accuracy is an essential quality of science writing and should not be violated to insure ease of comprehending the copy. Every effort should be made, however, to indicate the meanings of specialized terms.

Various reports were concerned with the application of readability formulas to materials at various levels. Andrews (1) studied the readability of Our finance Messenger, a Catholic classroom periodical for third grade pupils. The detailed steps taken to secure a valid appraisal are noteworthy. Mallison, Strum and Mallison applied the Flesch Ease of Reading formula to textbooks for high school physics (63) and for chemistry (64). Whereas the average difficulty of the texts examined was fairly satisfactory for use in the grades intended, wide variations in difficulty existed both among texts and with given texts which gave just cause for concern.

Siegel and Siegel (96) applied the Flesch readability formula to the pre-election speeches of Stevenson and Eisenhower. Whereas there were differences, the speeches of the two candidates did not differ significantly in difficulty level of interest appeal. Witherington (119) applied the Dale-Chall formula to eight textbooks in educational psychology and found that they varied in difficulty from grade ten to grade

twelve, the average being eleventh grade. Maloney (65) applied the Flesch Ease of Reading formula to Files How Supervise, which is intended for male industrial supervisors. It ranked at the high-school graduate level in difficulty. The bulletin was criticized because less than one-fourth of the industrial supervisors in the country are high-school graduates.

A new readability formula for primary grade materials was reported by Spache (101). After studying carefully the various factors that might be used in estimating the difficulty of such materials, the investigator followed "the example of Dale and Chall in choosing average sentence length as a predictive measure." In addition, use is made of the Dale list of 769 words, "judging all words outside this list as hard words." Needless to add that a readability formula especially designed for the primary level has been greatly needed. Forbes and Cottle (26) reported a new method for determining the readability of test passages based upon the weights assigned to words in the Thorndike Junior Century Dictionary.

Hygiene of reading. McNamara, Patterson and Tinker (61) secured data from 3,050 pupils to determine the influence of size of type on speed of reading. The type sizes used were 8, 10, 12, 14, 18, and 24 point. In grades I and II size of type "produced no consistent effect on speed of reading"; in grade III, 10, 12 and 14 point type was read faster than either 8, 18 or 24 point. It was concluded that speed of reading "should not be used as a criterion for choosing a type size for printing books to be used in the primary grades." Tinker (111) carried on a controlled experiment involving 285 college sophomores to determine the effect of intensity of illumination upon speed of reading six-point type. The illumination intensities used were 1, 10, 25 and 50 foot candle light. The data showed that speed of reading increased significantly with increase in candle power up to 25 with a further "slight but insignificant gain" as the intensity was raised to 50 foot candle. The investigator concluded that the critical level of illumination is between 10 and 25 with preference nearer 25 than 10. In still another study (112) Tinker found that speed of reading six-point type is reduced 5 per cent by vibration and that the combined influence of small type, vibration and use of italics produces a relatively large increase in speed of perception.

### The Teaching of Reading

Current status of reading achievement. Previous summaries in this series have reported many efforts to compare the efficiency of reading instruction today and formerly. All such evidence that could be secured has been reviewed again by Gray and Iverson (34). Two facts of large significance were emphasized. The first was that whereas

some progress has been made during the last three decades, particularly in understanding what is read, the progress in general has not been as great as might reasonably be expected. The second fact was that the data available are inadequate to insure valid comparisons.

A survey of the reading ability of pupils in the schools of Middlesborough, England (70) was carried on recently to determine the extent and causes of retardation in reading. Three tests were given to 2,236 pupils in the eleven and twelve age groups, namely the Burt Graded Word List Test, the Schonell Reading Comprehension Test and the Sleight Non-verbal Test of Intelligence. The scores made showed that 11 per cent of the pupils were 18 months or more retarded. Wide variations occurred among zones and schools, the poorest school having 23.6 per cent retardation. In addition to low mental ability the causes of retardation reported were "adverse home background and environment", "frequent or long absences", "specific defect-vision, hearing, etc.", and "frequent changes of school." Obviously the problems faced in England are similar in many respects to those faced in this country. As a means of coping more effectively with wide differences in the reading ability of pupils, Hamer (37) experimented with a plan of grouping them in a junior school in England on the basis of reading ability for purposes of reading instruction. The adoption of this plan resulted in greater progress than was normally expected and in the rapid reduction of the number of non-readers.

Relation of reading achievement to academic success. One reason for grave concern because of the low reading achievement of many pupils is the belief that they are definitely handicapped in other school subjects as a result. A study by Tilton (110) among fourth- and fifth-grade pupils led to the conclusion that ability to learn in one area is more or less consistent with "general ability to learn." The fact was emphasized, however, that failure to learn to read in harmony with expectation definitely affects progress in other areas. Studies by Barrett (7) by Pre and Botel (79) by Pitts (75) and by McQuarry (62) all provide evidence of more or less relationship between reading achievement and academic success. The findings support the conclusions reported in previous summaries that the extent of the relationship varies widely with the school subjects involved and that there are many factors other than ability to read that influence academic progress, such as general ability to learn, interests in specific fields, motives for taking given courses, etc.

Development of word recognition skills. Through the use of Section IV: Word Attack: Part 2, Silent, of the battery of Diagnostic Reading Tests, Triggs (114) made a study of the ability of readers from grade four through the college freshman year to hear and match letter sounds and to divide words into syllables. The test records showed some growth in ability to divide words into syllables but very little in ability

to hear and match letter sounds. Since the subjects did not in general approach the maximum score on the tests, such questions as the following arise. Was failure to improve due to the fact that they had acquired all the ability needed or did teaching fail to give the type of training needed?

Vocabulary studies. The Superior Educational Council of Puerto Rico, under the leadership of Dr. Rodriguez Bou (4) has prepared a Spanish word count in which serious effort was made to overcome the classical bias of previous word counts in that language. It was based on a total of 7,066,637 running words. Three sources of words were used: the vocabulary of expression of children; the vocabulary of recognition obtained from 91 issues of several newspapers; and the vocabulary proper for the elementary grades according to the individual judgment of authors of textbooks. It is believed that the list of words thus derived will be of great value in the preparation of readers and textbooks in Spanish and in guiding vocabulary mastery.

In order to develop a core vocabulary in the language arts of special value in teaching adult illiterates, Kyte (55) compiled a list of 663 words composed of the words in the first five hundred "of the Horn list of commonly written words (primarily from adult writings), and the first five hundred words in the Thorndike-Lorge list." This composite list was later checked against the most commonly used words of the Rinsland list. As a result the published list may be used advisedly in schools in determining the words that merit emphasis in teaching.

Meaning vocabulary problem. In order to study the extent and nature of the problems faced in teaching words of multiple meanings, Gammon (30) identified words used with various meanings in first, second, and third-grade readers. Tests were then prepared in which pupils checked pictures in different ways to indicate the meaning implied by the test directions. An analysis of pupil response showed that one-fourth of the answers of first-grade pupils, almost one-third of the answers of second-grade pupils, and over one-half of the answers of third-grade were incorrect. At each grade level there was a group of children, presumably having a richer background of experience, who made few, if any, incorrect responses. "Below this group the drop in errors was rapid." The errors were on a few words rather than "a smattering of errors for each word."

Effect of audio-visual aids on reading achievement. Two studies were concerned with the effect of the use of audio-visual aids in teaching reading upon achievement in reading. McCracken (59) experimented with the use of at least one film strip with every lesson in the basic reader used. These strips aimed "to stimulate pupil interest and to present the lesson it accompanied in a vivid and clarifying manner." As a result of evidence secured through the use of the film strips with different classes, the conclusion was reached that the use of such aids

is distinctly advantageous. Witty and Fitzwater (121) reported that the results of an experiment with a second-grade class in which the use of reading books that accompanied one set of eight films was used instead of the regular basic reader. The conclusion reached was that the use of films accompanying reading material provides needed basic experiences and "fosters the attainment of skills in reading." It not only leads to a clearer grasp of meaning but also enables the pupils "to hear the correct pronunciation of words and phrases in a meaningful setting."

Comparative effectiveness of listening and reading in incidental learning. Garland (31) compared the effectiveness of "listening to the radio" and "reading the newspaper" by means of a series of short tests of information based on four radio programs and four events of current interest reported in newspapers. Through the use of chi-squared tests it was found that differences in the results secured through listening and reading were not due to differences in the intelligence of the pupils. When those taking the tests were classified as "readers only", "listeners only" and "readers and listeners," the results were "generally favorable to the radio programme as a means of incidental learning, though the combination of listening and reading appears to be the best method." Stated differently those who both listened to radio programs and read newspapers tended to make the highest scores.

Interest appeal of pictures in texts. Increased attention has been given during recent years to the nature and effectiveness of non-verbal materials in texts. Whipple (117) analyzed the pictures in six fourth-grade texts to determine their significant characteristics and then carried on experiments with children to determine the interest appeal of each. Her findings were summarized in terms of the following standards which might be used as guides in estimating the interest appeal of the illustrations in books: "the proportion of pictures that have a definite central interest which draws the eye to a particular point"; "the proportion of the illustrations that depict action"; "the extent to which color is used"; "the average size of the illustration (the larger the area the more interesting)"; "the number of illustrations"; and "the extent to which the illustrations deal with eventful topics as opposed to still-life topics."

Map-reading ability of sixth-grade pupils. As a result of a careful study of pertinent literatures, including geography workbooks and teacher-aid suggestions, Wagner (115) prepared a detailed list of map-reading skills and knowledge. Tests were then prepared and administered to the pupils. The most significant conclusion reached was "that children do not have enough accurate knowledge in map-reading skills". The per cent of correct responses to a single item varied among individuals from 9 to 97. Evidence of growth in map-reading ability was

found in the fact that sixth-grade pupils ranked higher in skills introduced early in the grades than in those introduced later.

Comparative achievement of pupils in different language arts. Fea (24) reviewed previous studies and carried on an original investigation among fifth-and sixth-grade pupils to secure a clearer understanding of the interrelationships among speaking, reading, and writing. Nine different measures were used. For the specific group tested "the level of development in the three language arts appears to be in the order: material read, oral language and written language." Varying the order of the oral and written reproductions following the reading does not seem to effect their quality significantly. "However, there is a general trend toward superior language usage in both oral and written samples when written reproduction is performed first." In addition to these findings, almost twenty relationships between the three language arts were considered.

Methods of identifying reading needs and increasing the reading ability of high school students. Through the use of a questionnaire sent to 123 families, the staff of a midwestern junior high-school made a survey of the reading interests and habits of the pupils and their parents. In a summary of findings by Simpson and Camp (97) emphasis was given to the need of increased reading of non-fictional books and drama, "larger membership in book clubs", and increased use "of magazines rating those things consumers buy or use"; and "of acceptable magazines". The general procedure adopted was approved because of its value in helping to clarify teaching aims, to insure appropriate emphasis in teaching, and to direct and measure the improvement of learning in a community. As a result of an experiment with a high-school English class, Potter (77) secured evidence that through the proper stimulation, direction, and encouragement of reading the amount and quality of the personal reading of students could be definitely increased.

Effect of training in reading in the social studies. A means of developing the critical abilities required in the social studies to judge personal, social, and public issues intelligently, Brownell (7) gave two ninth-grade classes training in critical reading during a twenty-eight weeks period. The findings led to two recommendations that "significant gains in the more mature types of interpretation, critical reaction, reasoning and integration involved in efficient reading in the content fields can be made...as a result of a properly designed developmental reading program"; and that "significant gains in logical reasoning and critical judgment of written materials, as measured by the Watson-Glaser Critical Thinking Appraisal, or similar tests" can also be achieved.

Comparative value of three methods of teaching English. Evans (23) carried on an experiment for six months in the boys' departments

of six modern schools in the city of Birmingham, England, to determine the comparative value of the "project", "formal" and "oral" methods of teaching English. With respect to reading there was no significant difference in improvement ascribable to the methods used. The gains made correlated highly with the intelligence of the students. Improvement in word meaning and general knowledge proceeded best "if teaching consisted mainly of activities revolving around the spoken word". The results for the study as a whole support the hypothesis that the methods used should be adapted to the goals sought.

Corrective and remedial teaching. The problems faced by poor readers continue to elicit wide interest at all levels. Because of the increasing provision for remedial reading in England, Curr and Gourlay (13) carried on a study in primary schools to determine the effectiveness of two methods of selecting pupils, based on the results of the training provided. The results did not supply convincing evidence that selection of cases through the use of tests was superior to teacher selection. Differences in the "true gains" of the two groups were not statistically significant. The gains were greater on the standardized than on the unstandardized tests used. This was interpreted to mean that the training given aimed to develop reading skills involved in the standardized tests and had not transferred generally to other reading situations. Evidence secured by Preston (78) was not so favorable to teacher selection as compared with the use of tests in identifying of remedial cases. Basing judgments on the results of both reading and intelligence tests, he concluded that teachers tend to over-estimate the amount of retardation of some pupils and to fail to identify others who are retarded. He urged that greater attention be given to a child's mental ability in selecting remedial cases.

In an extended report on retarded readers in junior high schools by Lazar (56) a study of the intellectual capacity and functioning of 80 retarded readers was reported. Detailed analysis of their verbal and performance ~~finan~~s on the Wechsler-Bellevue Scales led to the conclusion that reading retardation is due in many cases to resistance in details rather than inability to learn. Counseling and intensive work with individuals was proposed as an appropriate approach to corrective or remedial training. Ephron (21) demonstrated the procedures and value of a continuous series of conferences with handicapped readers. As anxieties decrease they become free to make use of their intellectual ability and emotional energy in efforts to improve reading.

Coleman (11) presented evidence showing the value of what he called "a total push approach to remedial education" for many pupils of junior high-school age. It involved the individualization of instruction to meet the needs of each child with the following framework: "creation of a favorable learning atmosphere" and "of a need to learn"; "filling in weaknesses and gaps in the pupil's educational background"; "re-

medial work as an approach to the whole person"; "integration of home environment with school program"; use of "specific remedial techniques"; and "keeping the child up with his classes". In addition to a highly satisfactory "total mean improvement", favorable personality changes were noted. Delecato and Delecato (17) reported the results of a study which corroborated "the soundness of the rationale of the group approach to remedial reading" as reported a year ago. The program aimed to eliminate negative emotional reactions to reading and to the school and permitted choice of means by pupils in effecting improvement.

Nine studies relating to remedial reading at the college level reflect the rapid increase of interest during recent years in the reading needs of college students. Schubert (87) summarized answers to questions submitted to retarded and unselected college readers. The evidence showed clearly that the latter group differs from the former in significant personality traits, attitudes and study habits. In a second study, Schubert (88) found that retarded readers did not differ significantly in their hearing and reading vocabularies. Evidence that significant improvement can be made in the reading achievement of many poor readers at the college level was presented by Cutts (14) who provided vigorous emphasis on reading and study habits during a six-weeks' summer course before college entrance, and by Witty, Stolarz and Cooper (122) who provided varied types of training adapted to the needs of students during two nine-weeks' terms. The effect of remedial reading on academic grades was studied by Mouly (72) who found it to be effective in the case of all who made significant improvement in reading.

Johnson (47) compared the effect on reading improvement among college students "of a teaching technique which emphasized (1) drill on reading skills and (2) the psychology of learning with particular application to study skills." Whereas both methods increased reading skills to about the same extent the second resulted also in improvement on certain aspects of the ACE Psychological Examination and the California Test of Personality. Bennett (3) studied the effect of three methods on reading improvement in an English class: The use of the regular syllabus which outlined exercises in reading; freedom in reading, depending on intellectual curiosity and delight, with few requirements for outlining and written analysis, but with much oral analysis and critical discussion; and vigorous training in precise writing and more outlines, statements of theses, tests, and exercises of all types than the regular section received. Pre- and final-tests on the Diagnostic Reading Test showed that the third group made the best showing in all but vocabulary, which had been neglected purposely in that section; that the first group did the poorest in everything but speech; and that the second group did slightly better than the other groups in vocabulary. These findings support the hypothesis that different methods result in

different types of growth. Smith and Tate (98) presented evidence showing that the use of tachistoscopes and reading rate controllers resulted in substantial improvement in rate of reading. Cooper and Mills (12) supplied further evidence of the influence of the use of such mechanical devices on speed but pointed out the need for specific emphasis on comprehension when they are used.

Four studies were concerned with remedial training at the adult level. Staton and Maize (103) reported the results of five years of experimentation at the Air Command and Staff School at Maxwell Air Base, where training is provided largely through the use of rate controllers. The fact that the training resulted in improvement primarily in rate of reading was emphasized. It was believed, however, that this apparent improvement was due to "a change in reading habits and effort, rather than a true change in reading ability", as shown by the fact that speed of reading at the end of the training period was significantly less when taking a normal speed test than when reading with the rate controller. Jackson (44) made a subsequent report on remedial work at the Maxwell Air Base in which he emphasized the need of greater increase in both rate and comprehension to insure retention. Manolakes (68) compared the improvement made by an experimental group of marines who were given training through the use of both a tachistoscope and a rate controller and of a control group which used the rate controller only. "There were no significant differences between the groups indicated either prior to or following the training program". Holmes (40) presented evidence which showed that gifted adults can make substantial improvement in speed of reading through emphasis on "integrated teaching of the underlying skills."

Reading tests. Only a limited number of studies relating to reading tests was reported. Traxler (113) studied the scores made on the Diagnostic Reading Test: Survey Section, Lower Level, by pupils entering their responses in the test booklets and on separate sheets and found that the two groups did about equally well. As a result of statistical analyses it was concluded that this test is quite reliable for use in the intermediate grades. Frederiksen (28) found that separate timing of Parts I and II of the Cooperative English Test C2: Reading Comprehension produced "no significant changes in the means and standard deviations of the reading test scores, in the predictive value of the scores, their intercorrelations, or their correlation with other predictive measures". Swineford and Miller (107) found that the amount of guessing in answering test items was influenced by the kind of directions received, the group which was told not to guess giving answers to fewer items than the groups receiving no instructions about guessing. Hanes (38) found that the items on the Minnesota Multiphasic Personality Inventory which is "one of the most readable" of current tests were subject to numerous interpretations by subjects of limited education.

## BIBLIOGRAPHY

1. Andrews, Dorothy Irene, Geo. A. Pflaum Publishing Co., Dayton, Ohio. "A Study of the Vocabulary and Readability of a Third-Grade Classroom Periodical," Proceedings and Addresses Forty-Ninth Annual Meeting, Bulletin National Catholic Educational Association, (August, 1952), 221-225.  
Describes the steps taken, and the findings, in developing and evaluating a school paper for children, with special reference to third-grade pupils.
2. Barrett, Dorothy M. "Correlation of Survey Section of Diagnostic Reading Tests and of Test C2: Reading Comprehension With College History Grades," Journal of Educational Research, XLVI, (February, 1953), 465-9.  
Compares the coefficients of correlation between (a) the scores made by 200 college freshmen on the Survey Section of the Diagnostic Reading Test C2: Reading Comprehension and (b) an index of achievement in history derived from grades made during two terms in a history course taken by all freshmen.
3. Bennett, A. L. "An Experiment in Reading," Michigan Education Journal, XXX (January, 1953), 302-303.  
Compares the progress in vocabulary development and speed and comprehension in reading of groups taught differently: (a) free groups-wide reading and discussion; (b) regular sections—prescribed reading and quizzes; and (c) XP groups—brisk training in different aspects of reading except meaning vocabulary.
4. Bou, Ismael Rodriguez. "Spanish Word Count of the Superior Educational Council of the University of Puerto Rico," Journal of Educational Research, XLVI (April, 1953), 609-13.  
Describes the motives for preparing a Spanish word count and the three basic sources used (vocabulary of expression, vocabulary of recognition, vocabulary for the elementary grades according to individual judgments of authors of textbooks); reports the number of words derived from each source.
5. Bowman, Howard A. "Difference in Academic Achievement Between Pupils Who Left and Pupils Who Entered Los Angeles High Schools, 1948-1950," California Journal of Educational Research, III (November, 1952), 216-222.  
Based conclusions upon the results of the city-wide use of the Iowa Tests of Educational Development, including tests of understanding of basic social concepts, tests of interpretation of social studies, natural science and literary materials, and general vocabulary.
6. Brown, Clyde M. "Science Interests of Junior College Girls as

Determined by Their Readings in Current Science," Science Education, XXXVII (March, 1953), 105-108.

Analyzes 10,215 reading reports made by 217 Stephens College girls to determine (a) their areas of science interests, (b) the sources from which they read, and (c) the nature of a science course built upon their science interests.

7. Brownell, John Arnold. "The Influence of Training in Reading in the Social Studies on the Ability to Think Critically," California Journal of Educational Research, IV (January, 1953), 28-31.

Summarizes the results of an experiment conducted in two ninth-grade classes to determine the influence of two hours of training in reading for twenty-eight weeks on ability to think critically, as measured by the Watson-Glaser Critical Thinking Appraisal.

8. Burton, Dwight L. "The Relationship of Literary Appreciation to Certain Measurable Factors," Journal of Educational Psychology, XLIII (November, 1952), 436-39.

Analyzes the results of verbal and non-verbal intelligence tests, silent reading ability and studies of socio-economic status in the case of 190 twelfth-grade students to determine relationships, if any, with literary appreciation, as measured by the Carroll Prose Appreciation Test.

9. Clark, Willis. W. "Evaluating School Achievement in Basic Skills in Relation to Mental Ability," Journal of Educational Research, XLVI (November, 1952), 179-191.

Presents charts and tables showing variations in median achievement above or below the norms for the respective grades by pupils with varying median intelligence quotients.

10. Cofer, Charles N. and Shevitz, Reuben. "Word-Association as a Function of Word-Frequency," American Journal of Psychology, LXV (January, 1952), 75-79.

Reports the results of an experiment involving two groups of college students to determine "the relationship between word-frequency counts and number of associations given to stimulus-words of varying frequency."

11. Coleman, James C. "Results of a 'Total-Push' Approach to Remedial Education," Elementary School Journal, LIII, (April, 1953), 454-8.

Presents evidence of the value of an individualized integrated "bio-psychological approach" to educational retardation, including reading, in the case of ten boys and ten girls between the ages of eight and sixteen.

12. Cosper, Russell and Mills, Barriss. "Reading Comprehension and Speed," School and Society, LXXVII (June 6, 1953), 359-362.

Summarizes the gains made during given semesters among 700 college students as a result of "developmental training in reading"

- given in a two-hour a week laboratory course; considers the implications of the findings in relation to valid procedures in evaluation.
13. Curr, William and Gourlay, N. "An Experimental Evaluation of Remedial Education," British Journal of Educational Psychology, XXIII, Part I (February, 1953), 45-55.  
Reports the results of an experiment at the primary-school level to determine the relative effectiveness of two methods of selecting remedial cases and the value that can be obtained through remedial education.
14. Cutts, Warren G., Jr. "Dealing with Reading Problems at the Pre-College Level," Educational Administration and Supervision, XXXIX (March, 1953), 129-138.  
Describes the methods used and the results secured during a six-weeks' intensive reading program, half-day sessions, administered during the summer at Syracuse University.
15. Cutts, N. E., and Moseley, Nicholas. "Bright Children and the Curriculum," Educational Administration and Supervision, XXXIX (March, 1953), 168-173.  
Summarizes the responses of 673 pupils in grades II to XII inclusive, all having IQ's of 120 or above, to the questions: what school subject do you like best? Least? What subject makes you work hardest?
16. Deady, Marion C. "Visual Factors in Reading Disability," The Columbia Optometrist, XXVI (December, 1952), 5-7.  
Discusses the nature of visual difficulties that are related to reading retardation, as reported in eleven previous investigations; suggests means of correcting and eliminating each defect.
17. Delecato, Janice F., and Delecato, Carl H. "A Group Approach to Remedial Reading," Elementary English, XXX (January, 1953), 31-33.  
Presents evidence supporting the soundness of "the rationale of the group approach to remedial reading", as proposed by the same authors in the March 1952 issue of Elementary English.
18. Dunnette, Marvin D., and Maloney, Paul W. "Factorial Analysis of the Original and the Simplified Flesch Reading Ease Formulas," Journal of Applied Psychology, XXXVII (April, 1953), 107-110.  
Compares the amount of time and the accuracy of applying the original method of counting syllables and the simplified procedure, proposed by Farr, Jenkins and Paterson, in using the Flesch readability formula.
19. Edson, William H., Bond, Guy L., and Cook, Walter W. "Relationships Between Visual Characteristics and Specific Silent Reading Abilities," Journal of Educational Research, XLVI (February, 1953), 451-57.

- Reports the results of a study to determine the relationship, if any, between variations in ten measures of silent reading skills and thirteen tests of visual characteristics in the case of 188 fourth-grade pupils in four schools of St. Paul, Minnesota.
20. England, George W., Thomas, Margaret, and Paterson, Donald G. "Reliability of the Original and the Simplified Flesch Reading Formulas," Journal of Applied Psychology, XXXVII (April, 1953), 111-13.
- Bases conclusion on the performances of 13 pairs of relatively inexperienced analysts using samples drawn from house organs and a single more experienced analyst using samples drawn from books.
21. Ephron, Beulah Kanter. Emotional Difficulties in Reading: A Psychological Approach to Study Problems. New York: The Julian Press, Inc., 1953. Pp. xvi + 289.
- Reports a series of 39 interviews with a twenty-year old boy, and varying numbers of interviews with other youth, which were conducted to identify the nature of emotional blocks in reading and the progressive mastering of such difficulties through therapy.
22. Eppel, E. M., and Eppel, M., "Young Workers at a County College: A Pioneer Investigation of the Needs, Interests and Attitudes of 380 Young Workers Attending a County College." British Journal of Educational Psychology, XXIII (June, 1953), 87-96.
- Reports the results of inquiries concerning activities enjoyed most during leisure time and the amount devoted weekly to each, with special reference to "sports, cinema-going, dancing, reading, and club membership".
23. Evans, E. G. S. "An Experimental Study of Methods of Teaching English", British Journal of Educational Psychology, XXIII (June, 1953), 127-28.
- Comparative progress of boys from eleven to thirteen years of age attending six "modern schools" in "written expression", "reading ability", and "general knowledge" when taught for six months by one of three methods: "project", "formal", and "oral".
24. Fea, Henry R. "Interrelationships Among Materials Read, Written and Spoken by Pupils of the Fifth and Sixth Grades," Journal of Educational Psychology, XLIV (March, 1953), 159-75.
- Bases conclusion on a statistical treatment of data derived from analyses of oral-reading recordings, oral-reproduction recordings and written reproductions, in the case of 140 fifth- and sixth-grade pupils.
25. Figurel, J. Allen. "What Recent Research Tells Us about Differentiated Instruction in Reading," Reading Teacher, VI (September, 1952), 27-33, 44.

- Summarizes the results of related research under the headings: grouping of children, remedial instruction, reading readiness, and related studies.
26. Forbes, Fritz W., and Cottle, William C. "A New Method for Determining Readability of Standardized Tests," Journal of Applied Psychology, XXXVII (June, 1953), 185-90.  
Compares the grade placement of 27 standardized tests according to each of five readability formulas, the average of the five formulas, and the application of the new Forbes Formula to items in and instructions for these tests.
27. Forgays, Donald G. "The Development of Differential Word Recognition," Journal of Experimental Psychology, XLV, (March, 1953), 165-68.  
Summarizes the responses of twelve subjects at each educational level from grade two to ten inclusive and from grades thirteen to fifteen inclusive, to 20 three or four-letter English words, tachistoscopically presented to determine if educational training influences differential "recognition of words presented to the right and left of central fixation."
28. Frederiksen, Norman, "The Influence of Timing and Instructions on Cooperative Reading Test Scores," Educational and Psychological Measurements, XII (Winter, 1952), 598-607.  
Compares the effects "on means, standard deviations, inter-correlations and correlations with other measures, of two methods of timing of Parts I and II" and "of modified instructions for Part II" when the Cooperative English Test C2: Reading Comprehension (Higher Level), Form R, was given to two groups of entering college students.
29. Freeburne, Cecil Max. "A Study of the Relationship Between Figural After-Effect and Reading-Test Performance," Journal of Educational Psychology, LXIII (May, 1952), 309-12.  
Presents product-moment correlation coefficients between "the number of positive figural after-effects", the case of 24 psychology students, and their "subtest standard scores on the Iowa Silent Reading Test."
30. Gammon, Agnes L. "Comprehension of Words with Multiple Meanings," California Journal of Educational Research, III (November, 1952), 228-32.  
Presents the results of a study in which the words used with various meanings in readers for the first three grades were identified and tests prepared and given to determine the problems which children in each of these grades have in reading and understanding selected words.
31. Garland, G. R. "A Survey of the Extent to which School Children Listen to and Appreciate BBC Programmes and Its Relationship

to Their Acquisition of Knowledge," British Journal of Educational Psychology, XXII (November, 1952), 216-17.

Compares the scores made by school children on tests covering four radio programmes and four events of current interest reported in the newspapers to determine the relative effectiveness of radio listening and reading in the incidental acquisition of knowledge.

32. Graham, E. Ellis. "Wechsler-Bellevue and WISC Scattergrams of Unsuccessful Readers," Journal of Consulting Psychology, XVI (August, 1952), 268-71.

Compares the scores made by 96 unsuccessful readers, varying in ages from 8 years, 10 months, to 16 years, 11 months, on the various subtests of the WB and the WISC tests.

33. Gray, William S. "Summary of Reading Investigations July 1, 1951 to June 30, 1952," Journal of Educational Research, XLVI (February, 1953), 401-37.

Presents an annotated bibliography of 91 studies published during the year beginning July 1, 1951, and summarizes important findings under four headings: research in reading, the sociology of reading, the physiology and psychology of reading, the teaching of reading.

34. Gray, William S., and Iverson, William J. "What Should Be the Profession's Attitude Toward Lay Criticism of the Schools? With Special Reference to Reading," Elementary School Journal, LIII (September, 1952), 1-44.

Reviews evidence concerning the current status of reading instruction under such headings as: "Has reading been neglected"; Is as much time allotted to reading as formerly"; How well do pupils read"; "Are present levels of achievement high enough"; "How widely do pupils read"; "What is the quality of the material read"?

35. Grayum, ~~Heina~~ Stolte. "An Analytic Description of Skimming: Its Purpose and Place as an Ability in Reading," Studies in Education, 1952, pp. 137-143 (No. 4 in Thesis Abstract Series), Bloomington, Indiana: School of Education, Indiana University, 1953.

Discusses the methods and findings of an investigation which aimed "to discover the nature of skimming as an ability and its purpose and place in the curriculum by age-grade levels" (fourth grade to adults not in school).

36. Haigh, Gerard V. and Fiske, Donald W. "Corroboration of Personal Values as Selective Factors in Perception," Supplement to the Journal of Abnormal and Social Psychology, XLVII, (April, 1952), 394-98.

Reports the results of a study involving 27 university students

to determine the effect of preferences for 36 words upon time to recognize these words when presented tachistoscopically, the value preferences being measured both directly and indirectly.

37. Hamer, Clifford, "Reading in Junior School", The Times Educational Supplement, No. 1931 (May 9, 1952), 393.

Compares the progress made by pupils in a "junior school" in England classified in English on the basis of their reading ability with the amount of progress usually expected of pupils.

38. Hanes, Bernard. "Reading Ease and MMPI Results." Journal of Clinical Psychology, IX (January, 1953), 83-5.

Analyzes the scores of 100 male penitentiary inmates on the Iowa Silent Reading Test and their responses to the Minnesota Multiphasic Personality Inventory to test the hypothesis that this inventory "will communicate different amounts and not necessarily the identical information to subjects of limited reading ability."

39. Holmes, Jack A. "Gifted Adults Can Learn to Read Faster," California Journal of Educational Research, IV (May, 1953), 103-110.

Summarizes the results of an experiment with fifteen top executives of the Standard Oil Company of Ohio to determine the value of integrated teaching of the underlying skill factors in the reading process plus exercises in reading per se.

40. Holmes, Sidney, "The Selection and Use of Trade Books with Poor Readers," Clinical Studies in Reading, II, pp. 112-18. Supplementary Educational Monographs, No. 77, University of Chicago Press, 1953.

Reports the results of a study of the recreational reading interests of 17 remedial cases in reading, and of efforts to stimulate recreational reading during a period of ten weeks.

41. Huelsman, Charles B., Jr. "Educational Changes in an Experimental Classroom," Clinical Studies in Reading, II, 149-155. Supplementary Educational Monographs No. 77, University of Chicago Press, 1953.

Reports the results of an experiment involving fifth-grade pupils to determine the influence on achievement in reading of classroom changes, including illumination, to meet "Harmon's specifications."

42. Hughes, Mildred C. "Sex Differences in Reading Achievement in the Elementary Grades," Clinical Studies in Reading, II, pp. 102-106. Supplementary Educational Monographs No. 77. Chicago: University of Chicago Press, 1953.

Bases conclusions on an analysis of the scores made by pupils in grades III to VIII inclusive (an average of about 49 boys and 52 girls per grade) on the SRA Test of Primary Mental Abilities,

the Chicago Reading Tests, the Van Wagenen-Dvorak Rate Test, and the Huelsman Word Discrimination Test.

43. Hunt, Jacob Tate. "The Relation Among Vocabulary, Structural Analysis, and Reading," Journal of Educational Psychology, XLIV (April, 1953), 193-202.

Presents and interprets coefficients of correlation for 15 variables, based on the scores made by 168 college sophomores, juniors and seniors on tests given to determine relationships among vocabulary, structural analysis, and reading.

44. Jackson, James L. "A Progress Report on the Reading Laboratory at the Air Command and Staff School," School and Society, LXXVIII (July 11, 1953), 8-9.

Summarizes the results secured through remedial instruction at the air command and staff school and outlines the types of changes desirable in the program to secure improved results.

45. Jensen, Barry T. "Left-right Orientation in Profile Drawing," American Journal of Psychology, LXV (January, 1952), 80-83.

Presents data from four groups who read English, Norwegian and Arabic to determine if orientation in reading influences people in drawing profiles that face to the right or left.

46. Jensen, Barry T. "Reading Habits and Left-right Orientation in Profile Drawings by Japanese Children," American Journal of Psychology, LXV (April, 1952), 306-7.

Analyzes data from 8,856 Japanese school children ages 6-15 inclusive, to determine if the culturally imposed directional habits in reading and writing influence orientation in profile drawing.

47. Johnson, Granville B. "A Comparison of Two Techniques for the Improvement of Reading Skills at the College Level," Journal of Educational Research, XLVI (November, 1952), 193-205.

Compares the effects on the reading efficiency of college freshmen of two teaching procedures: (a) drill on reading skills, supplemented by discussions of the psychology of learning with particular application to study skills; and (b) group psychotherapy with emphasis on reading only as a part of the study of the psychology of learning.

48. Jones, Morris Val. "The Effect of Speech Training on Silent Reading Achievement," Journal of Speech and Hearing Disorders, XVI (September, 1951), 258-63.

Reports the results of a controlled experiment including 61 pairs of third-grade pupils from 10 classrooms to determine the effect of 36 speech improvement lessons upon progress in silent-reading achievement.

49. Jones, W. R. "The Influence of Reading Ability in English on the Intelligence Test Scores of Welsh-Speaking Children," British Journal of Educational Psychology, XXIII (June, 1953), 114-20.

Presents data, supplementing those previously reported, concerning the influence of reading ability on the scores made by Welsh-speaking (bilingual) children on an English intelligence test.

50. Jones, W. R. "The Language Handicap of Welsh-Speaking Children. A Study of Their Performance on an English Verbal Intelligence Test in Relation to Their Non-Verbal Mental Ability and Their Reading Ability in English," British Journal of Educational Psychology, XXII (June, 1952), 114-123.

Compares the performance of 117 Welsh-speaking children (bilingual) in five schools on a verbal intelligence test in English and a non-verbal test in Welsh and studies the differences found in relation to their reading ability in English.

51. Klare, George R. "Measures of the Readability of Written Communication: An Evaluation," Journal of Educational Psychology, XLIII (November, 1952), 385-99.

Presents the results of an evaluative study of three readability formulas — the Gray-Leary, Dale-Chall and the Flesch — as applied to 52 books.

52. Knox, Gertrude E. "Classroom Symptoms of Visual Efficiency," Clinical Studies in Reading, II, pp. 97-101. Supplementary Educational Monographs No. 77, Chicago: University of Chicago Press, 1953.

Presents evidence of the value of the classroom use of a check list of 30 visual characteristics as determined through its application to 126 third-grade pupils, the results of a visual screening test being reported in the case of 41 pupils and the findings of a refractionist in 37 cases.

53. Krise, Morely. "An Experimental Investigation of Theories of Reversals in Reading," Journal of Educational Psychology, XLIII (November, 1952), 408-22.

Summarizes data secured in a study involving 29 adult subjects "to investigate the degree of association, if any, between the tendency to commit reversals in reading and each of these factors": "lack of maturation"; "habitual direction of inspection"; "left-handedness"; "mixed dexterity"; "normal difficulties in space perception" and "lack of familiarity with the symbols involved."

54. Kruglov, Lorraine P. "Qualitative Differences in the Vocabulary Choices of Children as Revealed in a Multiple-Choice Test," Journal of Educational Psychology, XLIV (April, 1953), 229-43.

Compares the results of "a ten item multiple-choice vocabulary test in which three or four choices were correct but of different quantitative levels", when given to pupils "at the third-, fifth-, seventh- and eighth-grade levels, and to a group of college graduates."

55. Kyte, George C. "A Core Vocabulary in the Language Arts," Phi Delta Kappan, XXXIV (March, 1953), 231-4.  
Presents a list of 663 words which were derived by combining the 500 most frequently used words in the Horn basic writing vocabulary with the 500 most frequently used words in the Thorndike-Lorge list, and checked in terms of the 500 most commonly used words in the Rinsland list.
56. Lazar, May (editor). The Retarded Reader in the Junior High School: A Guide for Supervisors and Teachers. Bureau of Educational Research, Publication No. 31. New York City: Board of Education, 1952. Pp. 126.  
Shows the distribution of the reading-grade scores of 55,140 eighth-grade pupils on the Stanford Reading test; summarizes responses of 88 junior high-school principals concerning practices with retarded readers; presents results of a study of intellectual capacity and functioning of eighty junior high-school retarded readers.
57. Lepley, William M., and Kobrick, John L. "Word Usage and Synonym Representation in the English Language," Supplement to the Journal of Abnormal and Social Psychology, XLVII (April, 1952), 572-73.  
Reports the results of a study of the relationship between "number of synonyms and analogous words" given in Webster's Dictionary of Synonyms and "frequency of use", as given in The Teacher's Word Book of 30,000 Words.
58. Maclean, Malcolm S., Jr., and Hazard, William R. "Women's Interest in Pictures: The Badger Village Study." Journalism Quarterly, XXX (Spring, 1953), 139-162.  
Analyzes the responses of 152 women to 31 pictures in Time, U. S. News, and World Report, for 1949 and 1950, to identify common factors underlying women's interest in pictures; presents "correlations in interest for women between each picture and each group" and "intercorrelations among the six interest groups."
59. McCracken, Glenn. "The New Castle Reading Experiment - A Terminal Report," Elementary English, XXX (January, 1953), 13-21.  
Describes a series of experiments with first- and second-grade classes to determine the influence on progress in reading, as measured by the Gates Primary Reading Achievement Test of the use of film strip presentations before the book reading of each lesson.
60. McGinnies, Elliot, Comer, Patrick B., and Lacey, Oliver L. "Visual Recognition Thresholds as a Function of Word Length and Word Frequency," Journal of Experimental Psychology, XLIV (August, 1952), 65-69.

Reports the results of a study involving twenty subjects which attempted "to isolate the effects of both word length and word frequency upon thresholds of recognition, as measured by duration of exposure necessary for veridical report."

61. McNamara, Walter J., Paterson, D. G., and Tinker, Miles A. "The Influence of Size of Type on Speed of Reading in the Primary Grades," Sight-Saving Review, XXIII (Spring, 1953), 28-33, Bases conclusions on the scores made by 3,050 pupils when given three forms of a reading test printed in six different type sizes (8, 10, 12, 14, 18, 24.).
62. McQuary, John P. "Some Relationships Between Non-Intellectual Characteristics and Academic Achievement," Journal of Educational Psychology, XLIV (April, 1953), 215-28. Describes the results of a study "to determine the factor pattern underlying variables assumed to be related to scholastic achievement; presents correlations between some of these factors and such items as ACE Q, speed of reading, level of comprehension, vocabulary.
63. Mallinson, George Greisen, Sturm, Harold E., and Mallinson, Lois Marion, "The Reading Difficulty of Textbooks for High-School Physics," Science Education, XXXVI (February, 1952), 19-23. Reviews conclusions reached in previous studies of the reading difficulty of science textbooks; presents the readability scores of twelve high-school physics texts as determined through the use of the Flesch formula.
64. Mallinson, George Greisen, Sturm, Harold E., and Mallinson, Lois Marion. "Reading Difficulty of Textbooks for High-School Chemistry", Chemical Education, XXIX (December, 1952), 629-31. Discusses conclusions reached in previous studies concerning the difficulty of science textbooks; presents the readability scores of twenty high-school chemistry texts as determined through the use of the Flesch formula.
65. Maloney, Paul W. "Reading Ease Scores for File's How Supervise," Journal of Applied Psychology, XXXVI (August, 1952), 225-227. Compares the grade level of reading difficulty of File's, How Supervise, using the Flesch Readability Formula, with the grades reached in school by male industrial supervisors.
66. Malter, Morton S. "The Content of Current Comic Magazines," Elementary School Journal, LII (May, 1952), 505-510. Classifies the content of 185 comic magazines, published during the first two months of 1951, under five headings relating to "general layout" and eleven headings relating to "comic type" content.
67. Malter, Morton S. "Studies of the Effectiveness of Graphic Materials," Journal of Educational Research, XLVI (December, 1952), 263-73.

Summarizes major conclusions reached in ten studies of the relative effectiveness of various types of graphical material; includes suggestions concerning the grade placement of graphs.

68. Manolakes, George. "The Effect of Tachistoscope Training in an Adult Reading Program," Journal of Applied Psychology, XXXVI (December, 1952), 410-12.

Compares the progress in reading made by two groups of Marine Corps officers as a result of eighteen 25-minute sessions in which the experimental group substituted extended training in vocabulary and comprehension skills for tachistoscopic training given to the control group.

69. Maz, Veronica. "Radio Listening, Televiewing, and Reading Habits of Pittsburgh Public High-School Students," Pittsburgh Schools, XXVII (May-June, 1953), 156-72.

Summarizes 721 questionnaire replies providing data from 721 9a and 12a students and 1,224 parents and guardians.

70. Middlesbrough Education Committee. Report of a Survey of Reading Ability, Director of Education, Education Offices, Woodlands Road, Middlesbrough, England, 1953. Pp. 24.

Summarizes and interprets the scores on the Burt Graded Word List Test and the Schonell Reading Comprehension Test B given to 2,236 children between 11 and 12 years of age to determine the extent of backwardness in reading and the need for specific reading lessons in the lower forms of the secondary modern school.

71. Mitchell, June Frary, "Prediction of Increase in Silent Reading Rate," Clinical Studies in Reading, II pp. 89-93. Supplementary Educational Monographs No. 77. Chicago: University of Chicago Press, 1953.

Summarizes the results of an exploratory study including 84 cases ranging in chronological age from 12 to 20 and in IQ's from 92 to 125, to determine the value of a battery of texts, used by Rimoldi in a study of personal tempo, in the prediction of increase in rate of reading if proper instruction is given.

72. Mouly, George J. "A Study of the Effects of a Remedial Reading Program on Academic Grades at the College Level," Journal of Educational Psychology, XLIII (December, 1952), 459-66.

Compares the honor-points made during a period of two years by an experimental group of 155 students who took remedial reading for one semester with those of a control group of 164 students who were excused from such training for experimental purposes or who avoided taking the course in one way or another.

73. Mullaly, Sister Columba. The Retention and Recognition of Information. Educational Research Monographs, Vol. XVII, No. 3. Washington: Catholic University of America Press, 1952, Pp. xviii + 46.

Analyzes the results of a pretest, test and retest given at varying intervals to 1,278 pupils (in six schools) varying in ages from ten to fifteen years and in school grades from five through eight to determine the retention of information learned from one reading of a prose passage.

74. Perry, Dallis K. "Speed and Accuracy of Reading Arabic and Roman Numerals," Journal of Applied Psychology, XXXVI (October, 1952), 346-47.

Compares the reading performance of 30 college seniors and graduate students who were asked to read as rapidly and accurately as possible sets of numbers from 1 to 9, 10 to 49, and 50 to 99 in both Arabic and Roman numerals."

75. Pitts, Raymond J. "Relationship Between Functional Competence in Mathematics and Reading Grade Levels, Mental Ability, and Age," Journal of Educational Psychology, XLIII (December, 1952), 486-92.

Reports the results of a study involving 210 eleventh-grade girls in seventeen accredited Negro high schools in Georgia to determine the interrelationship of four variables: "functional competence in mathematics, reading grade levels, mental ability expressed in terms of Gamma IQ's, and chronological age."

76. Poling, Dorothy L. "Auditory Deficiencies of Poor Readers," Clinical Studies in Reading, II, pp. 107-111. Supplementary Educational Monographs, No. 77. Chicago: University of Chicago Press, 1953.

Analyzes the results of tests of auditory acuity, discrimination and memory span of 78 remedial cases in reading (58 boys and 20 girls) to determine whether deficiencies in these areas are related to specific errors in word discrimination.

77. Potter, Robert E. "Reading Unlimited," English Journal, XLII (January, 1953), 28-32.

Reports the results of an experiment involving 73 high-school pupils to determine the influence on amount and quality of providing a free period once each week for personal reading, of making interesting, attractive reading material available, and of avoiding the use of checking techniques which detract from the pleasure of reading.

78. Preston, Ralph C. "The Reading Status of Children Classified by Teachers as Retarded Readers," Elementary English, XXX (April, 1953), 225-7.

Presents and interprets the Wechsler IQ's and the reading indexes of 82 pupils in two schools who were "designated by their teachers as having distinct reading retardation."

79. Preston, Ralph C., and Botel, Morton. "The Relation of Reading Skill and Other Factors to the Academic Achievement of 2,048

College Students," Journal of Experimental Education, XX (June, 1952), 363-71.

Presents data showing the relation of achievement in reading, as measured by the Iowa Silent Reading Test, and college achievement, as measured by the means of the marks received by a student throughout his college career.

80. Reynolds, Maynard Clinton. "A Study of the Relationships Between Auditory Characteristics and Specific Silent Reading Abilities," Journal of Educational Research, XLVI (February, 1953), 439-49.

Presents the results of a statistical analysis of the scores made by 188 fourth-grade pupils on 14 measures of reading and auditory characteristics and specific silent reading abilities, as measured by the Gates Basic Reading Tests, types A and D.

81. Robinson, Helen M. (editor). Clinical Studies in Reading, II, Supplementary Educational Monographs, No. 77. Chicago: University of Chicago Press, 1953. Pp. x + 190.

Reports a series of studies of reading, with emphasis on visual problems, made by the staff members and graduate students doing work in the University of Chicago Reading Clinic.

82. Robinson, Helen M. "Diagnosis and Treatment of Poor Readers with Vision Problems." Clinical Studies in Reading, II, pp. 9-28. Supplementary Educational Monographs No. 77. Chicago: The University of Chicago Press, 1953.

Presents illustrative case studies of poor readers with visual problems, describes methods of visual screening, and discusses problems involved in remedial therapy.

83. Robinson, Helen M. "Personality and Reading," Modern Educational Problems (edited by Arthur E. Traxler), pp. 87-99. Report of the Seventeenth Educational Conference, New York City, October 30, 31, 1952, held under the Auspices of the Educational Records Bureau and the American Council on Education. Washington, D.C.: American Council on Education, 1953.

Summarizes the findings of research concerning personality and reading, grouping the studies reviewed in terms of the methods used in evaluating personality.

84. Robinson, Helen M., and Huelsman, Charles B., Jr. "Visual Efficiency and Progress in Learning to Read," Clinical Studies in Reading, II, pp. 31-63. Supplementary Educational Monographs No. 77. Chicago: University of Chicago Press, 1953.

Reports the results of studies, including more than 50 pupils in grades one, four and eight, to determine the relationship between visual efficiency and reading progress and to evaluate existing visual screening tests, determining their reliability and validity when used with elementary-school pupils varying in age and achievement.

85. Roesch, Raymond A. "Teaching Desirable Study Habits Through Experimentation," Catholic Education Review, LI (March, 1953), 152-61.  
Presents evidence of the effect on speed and comprehension in reading of practice exercises, which also provided records of progress in meaning vocabulary, speed of reading continuous text, and comprehension of New Testament chapters.
86. Russell, David H. "Interrelationships of the Language Arts and Personality," Elementary English, XXX (March, 1953), 167-90.  
Presents conclusions based on about 200 students concerning the general relationships between reading and personality and between reading difficulties and personality disturbances.
87. Schubert, Delwyn G. "A Comparative Study of Retarded and Unselected College Readers with Respect to Certain Study Habits, Attitudes and Personality Traits," Journal of Educational Research, XLVI (February, 1953), 471-4.  
Analyzes responses made by 50 retarded readers in college and 50 unselected college students to a list of 36 items on which they were asked to rate themselves and identifies items revealing significant differences between the two groups.
88. Schubert, Delwyn G. "A Comparative Study of the Hearing and Reading Vocabularies of Retarded College Readers," Journal of Educational Research, XLVI (March, 1953), 555-58.  
Compares the reading vocabulary, as measured by the Iowa Silent Reading Test, and the hearing vocabulary, as measured by the vocabulary section of form B.B of the California Reading Test, of 26 retarded readers in the reading clinic of Los Angeles State College.
89. Scipione, Alice M. "Eye-Movements as Related to Reading," The Columbia Optometrist, XXVII (February and May, 1953), 5-8; 3-4.  
Discusses significant characteristics of eye-movements in reading based on the results of 22 pertinent investigations.
90. Serra, Mary C. "How to Develop Concepts and Their Verbal Representations," Elementary School Journal, LIII (January, 1953), 275-85.  
Summarizes the findings reported in 34 studies concerning the influence of direct experience on concept development and vocabulary enlargement.
91. Serra, Mary C. "The Concept Burden of Instructional Materials," Elementary School Journal, LIII (May, 1953), 508-12.  
Summarizes the results of 14 studies relating to the concept burden of social studies material and basal reading materials.
92. Sharpe, Joseph Francis. The Retention of Meaningful Material. Educational Research Monographs, Vol. XVI, No. 8, June 1, 1952.

Washington, D. C.: The Catholic University of America, Pp. xii + 66.

Analyzes the nature and extent of the retention of once-read meaningful material, as revealed by a recall test administered to 2,492 children in grades five to ten inclusive, in 12 elementary and high schools, in an effort to find answers to controversial questions in this field.

93. Sheldon, William D., and Cutts, Warren C. "Relation of Parents' Home and Certain Developmental Characteristics to Children's Reading Ability, II," Elementary School Journal, LIII (May, 1953), 517-21.

Summarizes the information supplied by parents of 868 pupils in response to a questionnaire concerning ten characteristics of their children; compares findings with the status of the respective pupils in reading.

94. Shores, J. Harlan, and Saupe, J. L. "Reading for Problem-Solving in Science," Journal of Educational Psychology, XLIV (March, 1953), 149-58.

Analyzes the results of a "Test of Reading for problem solving in Science," given to 182 fourth, fifth, and sixth grade pupils to test the hypothesis that "reading ability differentiates beyond the primary grades into somewhat specific abilities to read different kinds of material for different purposes.

95. Shulman, Paul F. "The Vision Specialist in a Remedial Reading Program," Optometric Weekly, XLIII (1951), 1941-45.

Reviews the results of 20 studies relating to causation of reading disability to determine if vision tests alone can diagnose reading deficiency and if there is a definite syndrome of visual factors associated with reading difficulty.

96. Siegel, Arthur I., and Siegel, Estelle. "Flesch Readability Analysis of the Major Pre-election Speeches of Eisenhower and Stevenson," Journal of Applied Psychology, XXXVII (April, 1953), 105-6.

Compares the "reading ease" and "human interest" of published versions of speeches of the two presidential candidates, as determined through the use of the Flesch formula.

97. Simpson, Ray H., and Camp, Kenneth L. "Diagnosing Community Reading," School Review, LXI (February, 1953), 98-100.

Summarizes data concerning "out-of-school behavior goals in reading" secured the use of a questionnaire, filled out by the parents of secondary-school pupils in five schools and representing 95 families; points out potential values of such a survey.

98. Smith, Henry P., and Tate, Theodore R. "Improvements in Reading Rate and Comprehension of Subjects Training with the Tachistoscope," Journal of Educational Psychology, XLIV (March,

1953), 176-184.

Compares the progress in rate and comprehension in reading of 18 college students as a result of training given during 25 to 70 periods of 50 minutes each, half of each training period being spent in recognizing digits flashed on a screen and the other half in reading material on a reading rate controller.

99. Solomon, Richard L., and Postman, Leo. "Frequency of Usage as a Determinant of Recognition Thresholds for Words", Journal of Experimental Psychology, XLIII, (March, 1952), 195-201.

Criticizes the use of such indices of word frequency as the Thorndike list provides; reports the results of a controlled experiment involving university students as subjects to determine "the relationship between tachistoscopic recognition thresholds for words and frequency of prior usage of the words."

100. Solomon, Ruth H. "Personality Adjustment to Reading Success and Failure," Clinical Studies in Reading, II, pp. 64-82. Supplementary Educational Monographs No. 77. Chicago: University of Chicago Press, 1953.

Reports the conclusions reached through a study of the records of good and poor readers before and after the two years of instruction in reading to determine "the values of the Rorschach Test in investigating personality adjustment in relation to success and failure in reading.

101. Spache, George. "A New Readability Formula for Primary-Grade Reading Materials," Elementary School Journal, LIII (March, 1953), 410-13.

Reports correlations of "style elements and grade levels of textbooks" based on 224 samples of 100 words each from 152 books in common use in the first three grades; presents a formula for predicting readability of primary-grade reading materials based on the findings.

102. Spiegelman, Marvin, Terwilliger, Carl, and Tearing, Franklin. "The Content of Comics: Goals and Means of Comic Strip Characters," Journal of Social Psychology, XXXVII (May, 1953), 189-203.

Presents the results of a content analysis of 52 syndicated comic strips to determine "the goals sought and the goals approved or disapproved of comic strip characters and the means used in achieving them, noting differences between the sexes and socio-economic classes".

103. Staton, Thomas F., and Maize, Ray C. "A Voluntary Reading Improvement Program for Air Force Officers," School and Society, LXXVI (July 19, 1952), 42-44.

Describes the methods used and the results attained in the case of 1,832 officers through training in the reading of digits, letter

- words and continuous text, provided during 20 fifty-minute periods involving wide use of a reading-rate controller.
104. Stevenson, Lillian P., and Robinson, Helen M. "Eye-Hand Preference, Reversals, and Reading Progress," Clinical Studies in Reading, II, pp. 83-88. Supplementary Educational Monographs No. 77. Chicago: University of Chicago Press, 1953.  
Bases conclusions on the results of a study of the eye-hand preferences and directional tendencies of 60 kindergarten children and their subsequent reading progress over a period of approximately two years.
105. Sullenger, T. Earl, Parke, Libbie H., and Wallin, Willma K. "The Leisure Time Activities of Elementary School Children: A Survey of the Packing House Area of Omaha," Journal of Educational Research, XLVI (March, 1953), 552-4.  
Summarizes questionnaire returns from 1,800 children in grades four to eight of sixteen schools to determine the leisure-time activities in which they engaged, those they would enjoy but did not participate in, and reasons for not doing the things they would like to do.
106. Swanson, Charles E., and Fox, Harland G. "Validity of Readability Formulas," Journal of Applied Psychology, XXXVII (April, 1953), 114-118.  
Compares the responses relating to readership comprehension and retention, of 296 employees of a mid-western company to easier and harder versions of 12 articles in three issues of a monthly sent to employees.
107. Swineford, Frances, and Miller, Peter M. "Effects of Directions Regarding Guessing on Item Statistics of a Multiple-Choice Vocabulary Test," Journal of Educational Psychology, XLIV (March, 1953), 129-139.  
Summarizes and interprets data from 801 testees to determine "the amount of guessing that is likely to occur under different instructions... to find what relationship may exist between amount of guessing and performance in the area covered by the test, and to determine the effects of guessing on various statistics."
108. Tannenbaum, Percy H. "The Effect of Headlines on the Interpretation of News Stories," Journalism Quarterly, XXX (Spring, 1953), 189-197.  
Reports the results of an experiment involving 398 students to test two hypotheses: (1) "different headlines, presented with the same copy material, have different effects — that is give rise to different impressions"; (2) the effect "is in inverse proportion to the extent in which the story is read."
9. Tarleton, Raymond. "Accuracy and Comprehension in Science

- News Writing," Journalism Quarterly, XXX (Winter, 1953), 69-71.
- Compares the readability scores of 40 newspaper articles relating to science topics with the scientific accuracy of the articles to test the assumption that "inaccuracy in science writing increases when the ease of comprehension is increased."
110. Tilton, J. W. "The Intercorrelations Between Measures of School Learning," Journal of Psychology, XXXV, Second Half (April 1953), 169-79.
- Presents intercorrelations between the scores made by fourth-grade pupils on the various sections of the Unit Scales of Attainment, and between two testings on different forms at intervals of twenty months.
111. Tinker, Miles A. "The Effect of Intensity of Illumination Upon Speed of Reading Six-Point Italic Print." American Journal of Psychology, LXV (October, 1952), 600-2.
- Reports the results of an experiment involving 285 college sophomore, divided into a control and four experimental groups, to determine the effect on speed of reading six-point type with variations in the intensity of illumination (1, 10, 25, and 50 ft. c.)
112. Tinker, Miles A. "Effect of Vibration Upon Speed of Perception While Reading Six Point Print," Journal of Educational Research, XLVI (February, 1953), 461-4.
- Summarizes the results of experiments in a light laboratory with 25-foot candles of well-diffused illumination, involving four groups of students (one control and three experimental), to find out the effect of vibration upon speed in reading six-point type, as compared with 10 point type, with and without vibrations.
113. Traxler, Arthur E. "Results of the Diagnostic Reading Tests for Grades 4, 5 and 6, Survey Section, Among Independent School Pupils," 1952 Fall Testing Program in Independent Schools and Supplementary Studies, Educational Records Bulletin, No. 60, pp. 69-76. New York: Educational Records Bureau, 1952.
- Reports the results of studies concerning the validity of entering responses on separate answer sheets as contrasted with entering them in test booklets; also presents evidence concerning the reliability of the tests and intercorrelations among parts of the test and with other reading tests.
114. Triggs, Frances Oralind. "The Development of Measured Word Recognition Skills, Grade Four through the College Freshman Year," Educational and Psychological Measurement, XII (Autumn, 1952), 345-49.
- Compares scores made by testees in grades, four, five, six, ten and twelve, and the freshman year in college, on tests of ability to "hear and match" letter sounds; also compares the

scores of college students in different college departments to hear and match sounds and to divide words into syllables.

115. Wagner, Louise Durkee. "Measuring the Map-Reading Ability of Sixth-Grade Children," Elementary School Journal, LIII (February, 1953), 338-44.

Reports (a) the results of a survey of the map-reading abilities taught at the sixth-grade level, and (b) the per cent of correct answers of 100 sixth-grade pupils on test items dealing with specific map skills.

116. Werner, Heinz and Kaplan, Edith. The Acquisition of Word Meanings: A Developmental Study. Monographs of the Society for Research in Child Development, Inc., Vol. XV, Serial No. 51, No. 1. Child Development Publications of the Society for Research in Child Development, Inc. Evanston, Illinois: Fayerweather Hall, East, Northwestern University, 1950.

Summarizes the results of an experiment with children from nine to thirteen years of age which aimed to identify the processes underlying the acquisition of word meanings through verbal contexts.

117. Whipple, Gertrude. "Appraisal of the Interest Appeal of Illustrations," Elementary School Journal, LIII (January, 1953), 262-9.

Summarizes responses of 150 pupils of fourth-grade reading ability to six characteristics of pictures in fourth-grade textbooks (number of illustrations, size of pictures, color, center of interest, type of action, theme) with emphasis on their narrative interest appeal.

118. Williams, Gertrude H. "What Does Research Tell Us About Readiness for Beginning Reading?" The Reading Teacher, VI (May, 1953), 34-40.

Summarizes contributions of 33 studies relating to five aspects of readiness for beginning reading: "physical", "intellectual", "personal", "language", and "perceptual" readiness.

119. Withington, H. Carl. "Readability of Textbooks in Educational Psychology," Journal of Educational Psychology, XLVI (November, 1952), 227-230.

Compares eight textbooks in educational psychology published from 1948 to 1950 in respect to average length of sentence, percent of unfamiliar words, and raw scores secured by applying the Dale-Chall formula for predicting readability.

120. Witty, Paul. "A Study of Graduates of the 'Quiz Kids' Program," Educational Administration and Supervision, XXXVIII (May, 1952), 257-71.

Lists the titles of books read and enjoyed during six months by 41 adults who formerly had been selected and served on the "Quiz Kids" Radio-television programs; also the titles of books which

- in their judgment had had the greatest influence on them.
121. Witty, Paul, and Fitzwater, James P. "An Experiment with Films, Film-Readers, and the Magnetic Sound Track Projector," Elementary English, XXX (April, 1953), 232-41.  
Reports the average progress made by 27 second-grade children in each of six Chicago schools during a semester in which four films were presented, followed by the reading of "the associated film reader."
122. Witty, Paul, Stolars, Theodore, and Cooper, William. "Some Results of a Remedial Reading Program for College Students," School and Society, LXXVI (December 13, 1952), 376-80.  
Discusses reading retardation among college students, the procedure used at Northwestern University to identify students in need of help, the remedial program adopted, and the results attained.
123. Young, Frances A. "What People May Read," Educational and Administration Supervision, XXXVIII (May, 1952), 272-82.  
Presents in graphical form the number of new titles published yearly from 1880 to 1949; indicates by five-year periods the percent of the total number of titles that belong to each of 19 categories of the International Classification system.

# PHONIC KNOWLEDGE AND ITS RELATION TO THE SPELLING AND READING ACHIEVEMENT OF FOURTH GRADE PUPILS

MILDRED C. TEMPLIN

Institute of Child Welfare  
University of Minnesota

THERE IS no close agreement concerning the role of phonic training in the acquisition of reading and spelling skills although this problem has been discussed and studied over the years. However, there is some indication that children learn the more essential phonic principles whether or not formal instruction in phonics is given them, and that this knowledge is related to reading skill at the elementary grade levels. Tate (1), comparing the reading progress of first grade children given phonic drill with those trained in the "look and say" method, observed that children in both groups knew some fundamental phonic principles at the end of the training period. When Tiffin and McKinnis (3) investigated the relation between reading skill and the application of phonic principles to the pronunciation of nonsense-words by children in the fifth through the eighth grades, they did not consider the method by which the phonic principles were acquired. However, correlations ranging from .55 to .70 between reading skill and phonic scores suggest a substantial relation between knowledge of phonics and skill in reading.

The present study is concerned with the relation of phonic knowledge irrespective of how it was acquired, to spelling and reading achievement. It was designed to investigate: (a) the level of phonic knowledge of fourth grade pupils, (b) the relation between phonic knowledge and reading and spelling skill at this grade level, and (c) what differences, if any, exist in the phonic knowledge of good and poor spellers and of good and poor readers.

## The Tests Used

Phonic tests were constructed to determine the ability of children to associate sounds with the written symbols used to represent them. This sound-symbol association aspect of phonics was measured in one recall and three recognition tests. The ability of the children to discriminate sounds was measured by a sound discrimination test developed from previous work by the experimenter (2).

In the recall phonic test, the children wrote out the spelling of twenty-five different consonant sounds and sound combinations which were uttered by the experimenter. Throughout the study whenever a sound served as the stimulus, the sound of the letter and not the name of the letter

was presented. Although in most instances the children gave the most common spelling for the sound, credit was given for any spelling which occurs in English. For example, when the hard "c" was uttered, "k" and "c" were most frequently written, but those few children who wrote "ck" were, of course, given credit. One point for each correct response made the maximum recall score 25.

Three recognition measures of sound-symbol association were obtained in: (a) a word phonic test, (b) a nonsense-word phonic test, and (c) a sound phonic test. All were written multiple choice tests for which appropriate blanks were furnished the children.

The word phonic test used a familiar word uttered by the experimenter as the stimulus for each item. The subjects listened for a sound in a particular position in the word, and then selected from four letters that letter which represented the specific sound to which they had listened. For example, when the beginning "b" was to be identified, the experimenter said the word "boy" and the choice of the letter which represented that sound was made from "d," "g," "b," and "p." In the stimulus words for the various test items, initial consonants, initial consonant blends, final consonants, vowels and diphthongs were identified. Whenever possible, the sounds from which the choices were to be made were selected systematically to include the correct sound, the voiced or voiceless equivalent of this sound, and other sounds of the same type. In the example above, the correct sound was a voiced plosive and the distractors were "p," its voiceless equivalent, and "d" and "g," both voiced plosives. To be certain that the children understood the task, an example was presented whenever sounds of different types or in different positions in the word were to be identified. All of the stimulus words were familiar to the children. They were taken either from the easiest words on the Thorndike Word List or from the Wauwatosa Word List.<sup>1</sup> The test consisted of 83 items: 24 in which the beginning consonants were identified, 27 beginning blends, 20 final consonants, and 12 vowels and diphthongs. Since one point was given for each correct response, the total possible score was 83.

With the nonsense-word phonic test the multiple-choice test blank for the word phonic test was used, but the stimuli were nonsense-words. In the example given above, while the choice was to be made from the same letters, the stimulus word was "bakta" instead of "boy." The maximum score was again 83.

In the sound phonic test the children listened to a stimulus sound uttered by the experimenter and selected from three or four words the

1. The list is made up of the words used in the preprimers, primers, and first readers of five primary reading series.

one in which that sound appeared in the beginning or at the end. In most instances, a consonant sound was presented in isolation. Only when the vowel in the first syllable of each choice word was identical was the vowel uttered with the consonant sound. The experimenter said the syllable "fi" when the initial "f" sound was to be identified in the words "hide," "side," "find," and "pine"; but "sh" was uttered in isolation when the choice was to be made between "show," "sit," "children," and "thank." A total of 25 initial consonants, 29 initial consonant blends, 20 final consonants, and 14 vowels and diphthongs were identified in words. The maximum score became 88 since one point was allowed for each correct response.

A sound discrimination test measuring ability to discriminate between consonant sounds was presented orally by the experimenter. The items making up this test were the fifty most discriminating items from a 200 item test of sound discrimination.<sup>2</sup> In each of the fifty pairs of items the discriminating element was the consonant while the vowel was the same. The consonant appeared at the beginning, in the middle, or at the end of the items; e.g., "pe-be," "ewe-ele," "es-ef." On the test blank an "S" and a "D" appears after each number from one to fifty. After the experimenter said each pair of items, the children encircled the "S" if the items in the pair were the same and the "D" if they were different. One point was given for each correct discrimination with a resulting total possible score of 50.

As measures of spelling and reading ability, the children were given the spelling test from the Intermediate Stanford Achievement Test and the Durrell-Sullivan Reading Test.

#### Sampling

The sample consisted of 318 children from five Minneapolis public schools who were in grade 4.9 at the time the testing was carried on. Only those subjects were included in the analysis who had normal hearing, for whom intelligence test scores were available, and who had taken all of the phonic tests together with the standardized reading and spelling tests. Fifty-eight cases were dropped because of incomplete data.

Normal hearing was determined by a sweep check audiometric test which is part of the school testing program. Some measure of intelligence was available on all subjects. With few exceptions this was a Kuhlman-Anderson IQ obtained shortly after the child's entry into fourth grade. The mental age used for each child was corrected to the date of administration of the phonic tests.

2. These fifty items have not been previously published.

The children ranged in CA from 8 years 8 months to 12 years 0 months, with a mean CA of 10 years 10 months. Mental ages ranged from 8 years 0 months to 12 years 6 months with a mean of 10 years 3 months. On the spelling section of the Intermediate Stanford Achievement Test the mean was 32.4 and the median was 34.8. The median normative score for grade 4.9 is 36. On the Durrell-Sullivan Reading Test the mean was grade 6.0 and the median was grade 5.9. The experimental sample was accelerated in reading achievement and slightly below average in spelling achievement.

#### Administration of the Tests

The tests were given in classroom situations over a two week period. Classroom teachers gave the reading test as part of the school testing program. The spelling and phonic tests were administered by the experimenter in two or three sessions, whichever best fit into the particular school program. In no instance were the word phonic and the non-sense-word phonic tests presented on the same day. The recall phonic test was always the last to be administered.

In all of the phonic tests, the experimenter stood in the middle of the classroom facing the front of the room so that none of the children could observe her lip movements. This position reduced the distance from the experimenter to any particular child and equalized, as much as possible, the distance from the experimenter to all of the children.

#### Analysis of the Data

The scores of the entire sample were analyzed to determine the level of phonic knowledge and the interrelationship between the scores. The relation of phonic knowledge to spelling and reading achievement was further studied in contrasting groups of good and poor spellers and good and poor readers. Since the correlations of MA with the scores on the spelling and reading tests were substantial (.54 on the reading and .62 on the spelling tests) intelligence was controlled in the contrasted groups. To obtain cases which would be at the typical mental age for grade 4.9, those cases whose mental ages fell within three months of 9 years 9 months, the average CA for children in grade 4.9 were selected. The 86 cases in this MA range had a mean CA of 10 years 0 months and a mean IQ of 98. Using the scores on the Stanford Spelling Test to select the deviate spelling groups, the one-third receiving the lowest scores formed the lower spelling group and the one-third receiving the highest scores the upper. There were 26 cases in each of the deviate groups. The same technique was used with the scores on the Durrell-Sullivan Reading Test to select the upper and the lower reading cases.

Since the correlation between spelling and reading was substantial (.70 for the total sample) there was some overlap in the cases included in the deviate spelling and reading groups. Of the cases selected to form these contrasting groups, 16 cases appear in both the lower spelling and lower reading groups; 13 in the upper spelling and upper reading; four in the lower reading and upper spelling; and three in the upper reading and lower spelling.

### Results

Analysis of the Total Sample. —In Table I the mean raw and percentage scores on the sound discrimination and recall and recognition phonic tests are presented. There is a considerable range in scores for the recall and recognition phonic tests, and only on the sound discrimination test does a piling up occur at the maximum possible scores. These tests are comparatively easy for fourth-grade pupils.

The highest mean percentage scores are obtained on the sound discrimination and the word phonic test. A lower percentage score is made on the recall phonic than on any of the recognition phonic tests. This finding is in agreement with results of achievement testing in other areas when both recall and recognition measures are taken.

Phonic knowledge is not applied equally in all of the recognition tests. It is applied most successfully when a sound is to be associated with a symbol in a familiar word, and least successfully when a sound is to be associated with a symbol in a nonsense-word. A difference of over 10 percentage points in the scores is obtained. As the tests were designed, the word phonic test measures the application of phonic knowledge in familiar words and the nonsense-word phonic test measures the application in new words. Since nonsense-words were used, the unfamiliarity of the stimulus to all subjects was assured. The sound phonic test presents a somewhat novel situation since the sound rather than the word serves as the stimulus.

In order to determine the statistical significance among the differences in scores, the CR of the difference was determined between the sound discrimination test and each of the phonic tests, and among the several phonic tests. The differences in the mean percentage scores on these various types of phonic tests are significant at the .01 level of confidence in all instances except between the scores on the sound discrimination and word phonic tests. These are the tests on which the highest percentage scores were obtained. The difference between these scores is only .78 of one percent with a CR of 1.30. The other CR's range from 6.10 between the recall and nonsense-word phonic tests to 29.2 between the recall and both the word phonic and the sound discrimination measures. For the experimental sample there are real differences in the degree of application of phonic knowledge by fourth grade

The children ranged in CA from 8 years 8 months to 12 years 0 months, with a mean CA of 10 years 10 months. Mental ages ranged from 8 years 0 months to 12 years 6 months with a mean of 10 years 3 months. On the spelling section of the Intermediate Stanford Achievement Test the mean was 32.4 and the median was 34.8. The median normative score for grade 4.9 is 36. On the Durrell-Sullivan Reading Test the mean was grade 6.0 and the median was grade 5.9. The experimental sample was accelerated in reading achievement and slightly below average in spelling achievement.

### Administration of the Tests

The tests were given in classroom situations over a two week period. Classroom teachers gave the reading test as part of the school testing program. The spelling and phonic tests were administered by the experimenter in two or three sessions, whichever best fit into the particular school program. In no instance were the word phonic and the non-sense-word phonic tests presented on the same day. The recall phonic test was always the last to be administered.

In all of the phonic tests, the experimenter stood in the middle of the classroom facing the front of the room so that none of the children could observe her lip movements. This position reduced the distance from the experimenter to any particular child and equalized, as much as possible, the distance from the experimenter to all of the children.

### Analysis of the Data

The scores of the entire sample were analyzed to determine the level of phonic knowledge and the interrelationship between the scores. The relation of phonic knowledge to spelling and reading achievement was further studied in contrasting groups of good and poor spellers and good and poor readers. Since the correlations of MA with the scores on the spelling and reading tests were substantial (.54 on the reading and .62 on the spelling tests) intelligence was controlled in the contrasted groups. To obtain cases which would be at the typical mental age for grade 4.9, those cases whose mental ages fell within three months of 9 years 9 months, the average CA for children in grade 4.9 were selected. The 86 cases in this MA range had a mean CA of 10 years 0 months and a mean IQ of 98. Using the scores on the Stanford Spelling Test to select the deviate spelling groups, the one-third receiving the lowest scores formed the lower spelling group and the one-third receiving the highest scores the upper. There were 26 cases in each of the deviate groups. The same technique was used with the scores on the Durrell-Sullivan Reading Test to select the upper and the lower reading cases.

Since the correlation between spelling and reading was substantial (.70 for the total sample) there was some overlap in the cases included in the deviate spelling and reading groups. Of the cases selected to form these contrasting groups, 16 cases appear in both the lower spelling and lower reading groups; 13 in the upper spelling and upper reading; four in the lower reading and upper spelling; and three in the upper reading and lower spelling.

### Results

Analysis of the Total Sample. —In Table I the mean raw and percentage scores on the sound discrimination and recall and recognition phonic tests are presented. There is a considerable range in scores for the recall and recognition phonic tests, and only on the sound discrimination test does a piling up occur at the maximum possible scores. These tests are comparatively easy for fourth-grade pupils.

The highest mean percentage scores are obtained on the sound discrimination and the word phonic test. A lower percentage score is made on the recall phonic than on any of the recognition phonic tests. This finding is in agreement with results of achievement testing in other areas when both recall and recognition measures are taken.

Phonic knowledge is not applied equally in all of the recognition tests. It is applied most successfully when a sound is to be associated with a symbol in a familiar word, and least successfully when a sound is to be associated with a symbol in a nonsense-word. A difference of over 10 percentage points in the scores is obtained. As the tests were designed, the word phonic test measures the application of phonic knowledge in familiar words and the nonsense-word phonic test measures the application in new words. Since nonsense-words were used, the unfamiliarity of the stimulus to all subjects was assured. The sound phonic test presents a somewhat novel situation since the sound rather than the word serves as the stimulus.

In order to determine the statistical significance among the differences in scores, the CR of the difference was determined between the sound discrimination test and each of the phonic tests, and among the several phonic tests. The differences in the mean percentage scores on these various types of phonic tests are significant at the .01 level of confidence in all instances except between the scores on the sound discrimination and word phonic tests. These are the tests on which the highest percentage scores were obtained. The difference between these scores is only .78 of one percent with a CR of 1.30. The other CR's range from 6.10 between the recall and nonsense-word phonic tests to 29.2 between the recall and both the word phonic and the sound discrimination measures. For the experimental sample there are real differences in the degree of application of phonic knowledge by fourth grade

TABLE I

MEAN RAW AND PERCENT SCORES ON PHONIC TESTS  
FOR TOTAL SAMPLE

Test	Raw Score		Percent Score	
	N	S. D.	N	S. D.
Recall	18.83	1.42	79.09	5.97
Recognition:				
Word	77.67	4.94	93.20	5.93
Nonsense-Word	68.96	7.93	82.75	8.92
Sound	73.75	5.11	88.50	6.14
Sound Discrimination	46.99	4.58	93.98	9.16

TABLE II

CORRELATION OF SCORES ON PHONIC TESTS WITH  
MENTAL AGE, SPELLING AND READING

Test	M. A.	Spelling	Reading
Recall	.24	.34	.25
Recognition:			
Word	.43	.54	.40
Nonsense-Word	.36	.55	.44
Sound	.37	.57	.47
Sound Discrimination	.29	.23	.22

TABLE IV

SIGNIFICANCE OF DIFFERENCES ON MA, READING, SPELLING,  
 AND PHONIC SCORES FOR UPPER AND LOWER  
 DEVIATE READING GROUPS

Test	Mean		$\sigma$ diff.	C. R.	Level of Significance
	Upper (N = 26)	Lower (N = 26)			
M. A.	118.12	117.54	.54	1.06	....
Spelling	83.08	53.54	2.64	11.23	.001
Reading	34.54	24.38	2.05	4.96	.001
Recall	18.50	17.85	.88	.74	....
Recognition:					
Word	77.69	75.96	1.41	1.16	....
Nonsense-Word	70.31	65.96	2.18	1.99	.20
Sound	78.08	74.23	1.51	2.43	.05
Sound Discrimination	46.69	46.50	1.17	.16	....

pupils under various conditions.

The relation of scores on the phonic tests with MA, spelling, and reading is presented in Table II. The correlation between phonic test scores with spelling achievement is higher than with reading achievement on all except the sound discrimination test. A somewhat higher relationship to reading, spelling, and MA is found with each of the recognition phonic tests than with the recall phonic test.

In agreement with other investigators, the correlation between intelligence and achievement in reading and spelling is substantial and is greater with reading than with spelling. The correlation of MA with reading is .62, with spelling .54. The correlation between spelling and reading scores for this group is .70. The correlation between CA and reading is -.29 and with spelling -.18.

As Tate had observed with first grade subjects, these fourth graders had attained a substantial amount of phonic knowledge. Although all principals and teachers were queried about the phonic training of the classes included in the study, no reliable estimate of the differences in the amount of phonic experience was possible. The teachers' reports concerned only the fourth grade experience of the subjects, and the estimates of the teachers and principals even for this present grade were not in close agreement. Most of the teachers reported that they used phonic training when they felt it was necessary.

Analyses of the Contrasted Spelling and Reading Group. —The background factors and the phonic test scores for the upper and lower deviate spelling groups are presented in Table III. There is no significant difference in the mental ability of the two groups. The difference in the spelling score is certain, and a real difference also exists in reading skill.

On the recall and recognition phonic tests the better spellers receive the higher total score. The difference in the mean scores obtained on the sound discrimination test is negligible. The word phonic test, where the stimuli are familiar words, the difference between the deviate groups in ability to apply phonic knowledge is not significant. On the nonsense-word phonic test, where application of phonic principles is made in "new" words, the better spellers are more successful than the poor spellers although the difference is significant at only the twenty percent level of confidence. In the sound phonic test, when a sound is presented as a stimulus the upper deviate spellers are able to apply their phonic knowledge better at the two percent level of confidence. This situation is probably the most unfamiliar of the three recognition test situations to fourth graders.

It was considered that the greater difference observed on the sound phonic test might be related to differences in ability to hear the stimulus sounds produced in the classroom. However, this is unlikely, since in the recall and sound discrimination tests, sounds also serve as the

TABLE III

SIGNIFICANCE OF DIFFERENCES ON MA, READING, SPELLING,  
AND PHONIC SCORES FOR UPPER AND LOWER  
DEVIATE SPELLING GROUPS

Test	Mean		$\sigma$ diff.	C. R.	Level of Significance
	Upper (N = 26)	Lower (N = 26)			
M. A.	117.96	117.41	.55	1.00	....
Spelling	38.39	20.85	1.43	12.26	.001
Reading	76.16	59.43	3.99	4.19	.001
Recall	18.46	17.27	.74	1.61	.20
Recognition:					
Word	77.77	76.08	1.42	1.19	....
Nonsense-Word	69.39	65.85	1.96	1.81	.20
Sound	78.85	73.92	1.90	2.60	.02
Sound Discrimination	46.23	46.96	1.22	0.59	....

TABLE V

SIGNIFICANCE OF DIFFERENCES BETWEEN PERCENT OF POSSIBLE SCORES ON  
RECALL AND RECOGNITION PHONIC TESTS FOR UPPER AND LOWER DEVIATE  
SPELLING AND READING GROUPS

Test	Mean	Diff.	$\sigma$ Diff.	C. R.	Level of Significance
<u>Upper Reading Group (N = 26)</u>					
Recall	74.00				
Recognition:					
Word	93.62	19.62	2.03	9.67	.001
Nonsense-Word	84.72	10.72	2.44	4.39	.001
Sound	88.70	14.70	2.04	7.61	.001
<u>Lower Reading Group (N = 26)</u>					
Recall	71.40				
Recognition:					
Word	91.53	20.13	3.33	6.05	.001
Nonsense-Word	79.48	8.08	3.66	2.21	.05
Sound	84.32	12.92	3.36	3.81	.001
<u>Upper Spelling Group (N = 26)</u>					
Recall	73.84				
Recognition:					
Word	93.71	19.87	1.84	10.80	.001
Nonsense-Word	83.61	9.77	2.01	4.86	.001
Sound	89.57	15.73	1.97	8.03	.001
<u>Lower Spelling Group (N = 26)</u>					
Recall	69.08				
Recognition:					
Word	91.68	22.60	2.87	7.87	.001
Nonsense-Word	79.35	10.27	3.03	3.39	.01
Sound	83.97	14.89	2.87	5.19	.001

TABLE VI

SIGNIFICANCE OF THE DIFFERENCE BETWEEN PERCENT OF POSSIBLE SCORES ON  
RECOGNITION PHONIC TESTS\*

	N	Percent Means			Differences Between 1 - 2		Differences Between 1 - 3		Differences Between 2 - 3	
		Word Test (1)	Nonsense Word Test (2)	Sound Test (3)	$\sigma$ Diff.	C. R.	$\sigma$ Diff.	C. R.	$\sigma$ Diff.	C. R.
Total Sample	318	93.20	82.75	88.50	1.07	9.78	0.85	5.53	1.08	5.32
Upper Reading Deviates	26	93.62	84.72	88.70	1.69	5.27	1.04	4.73	1.70	2.34
Lower Reading Deviates	26	91.53	79.48	84.32	2.64	4.56	2.24	3.22	2.51	1.93
Upper Spelling Deviates	26	93.71	83.61	89.57	1.64	6.16	1.22	3.39	1.77	3.37
Lower Spelling Deviates	26	91.68	79.35	83.97	2.41	5.12	2.19	3.52	2.41	1.92

\*For the deviate groups a C. R. of 3.73 is significant at the .001 level; 2.79 at the .01 level; and 2.06 at the .05 level.

stimuli. The difference reaches only the twenty percent level of confidence in the recall test and the higher score is obtained by the poorer spellers in the sound discrimination test. It is more probable that the high deviate spellers are better able to apply their phonic knowledge and adapt themselves more readily to this somewhat unusual task.

Table IV presents background factors and the phonic test scores of the upper and lower deviate reading groups within the restricted mental age range. There is no significant difference between the mental age of the two groups of readers, although the better readers have the slight advantage of about one-half month of mental age. The differences in the test scores on both spelling and reading are certain, although the greatest difference occurs in the reading test scores. All of the phonic test scores are higher for the better readers. The differences between the two groups reach the five percent level of confidence only when the sound is used as a stimulus.

The difference in the scores on the word phonic test is not significant between the deviate spelling groups or between the deviate reading groups. The difference between the scores on the nonsense-word phonic test reaches only the twenty percent level of confidence between the deviates in the spelling group and in the reading group. On the sound phonic test and in the recall phonic test the level of significance of the difference is greater for the spelling than for the reading groups. Differences in sound discrimination ability are not significant for either sample. Although there is some overlap of cases in the spelling and reading deviate groups, it is likely that the relation of phonic knowledge with spelling is somewhat greater than with reading since the correlations for the experimental sample are higher between spelling and phonic scores than between reading and phonic scores.

When the phonic test scores are reported as percentage of possible score, the recall test scores are consistently lower than the recognition scores obtained for all contrasted groups. As seen in Table V this difference reaches the five percent level of confidence or higher for all of the contrasted groups. This is in agreement with the findings for the total sample.

The significance of the differences between the percentage of possible scores on the three recognition tests is presented in Table VI. For all deviate groups, as for the total sample, the highest scores are obtained on the word phonic test, the next highest on the sound phonic test, and the lowest on the nonsense-word phonic test. Differences for all groups between the percentage of possible scores on the word phonic and the nonsense-word phonic tests and on the word and sound phonic tests are significant at the one percent level. The differences between the non-sense-word and sound phonic tests are significant for the good spellers and readers at least at the five percent level, but below this level for the poor spellers and poor readers. For these poor spellers and read-

ers the application of phonic knowledge in both the "new" and the unfamiliar sound situations present real difficulties. The percentage scores indicate that both of these tasks are more difficult for the lower deviates in spelling and reading than the application of phonic principles when a more familiar word is the stimulus.

### Conclusions

1. A substantial amount of phonic knowledge as measured by the sound discrimination and sound-symbol association tests has been acquired by fourth grade pupils.
2. For the total sample, the correlations between phonic knowledge and spelling are somewhat higher than between phonic knowledge and reading.
3. A comparison of the scores of the deviate spelling groups indicate that (a) the better spellers receive the higher scores on all tests except the sound discrimination test; and (b) the differences in scores reach the .02 level of confidence for the sound phonic recognition test, and are below this level on the word phonic recognition and the sound discrimination tests.
4. A comparison of the scores of the deviate reading groups indicate that (a) the better readers receive the higher scores on all tests; and (b) the differences in scores reach the .05 level on the sound phonic recognition test, the .20 level for the nonsense-word phonic recognition test, and are below this level for the sound discrimination, recall and word phonic recognition tests.
5. Scores on the test of phonic knowledge are significantly higher when a recognition rather than a recall technique of measurement is used. This holds for the total sample and for each of the deviate spelling and reading groups.
6. Among the recognition measures, significantly higher scores are obtained when the stimulus is a familiar word rather than a sound or a nonsense-word. The differences among the scores on the tests using three different stimuli are significant above the .05 level for the total sample, the good spelling and good reading deviate groups. For the poor spelling and poor reading deviates, however, the differences between the word phonic and the other recognition tests are significant, but the differences between the nonsense-word and sound phonic tests are not.

That the poor spellers and poor readers applied their phonic knowledge less well than good spellers and good readers in the unfamiliar test situations while the difference was not significant when phonic knowledge was measured in familiar words is an intriguing finding. The degree of understanding of sound-symbol association differs for the upper and lower deviate groups although the measured scores show little difference in the familiar test situation. This may indicate a real difference in the ability of children of similar intellectual level to transfer what they know from one situation to another. On the other hand it may be related to the various methods of teaching used with these children or to the particular testing procedure used in this study. As the present investigation was designed, a conclusive answer cannot be given. However, the implication of this finding is important enough for educational procedure and psychological theory to warrant thorough and controlled investigation.

#### REFERENCES

1. Tate, H. L. "Influence of Phonics on Silent Reading in Grade I," Elementary School Journal, XXXVII (1937), pp. 352-363.
2. Templin, Mildred C. "A Study of Sound Discrimination Ability of Elementary School Pupils," Journal of Speech Disorders, VIII (1943), pp. 127-132.
3. Tiffin, J., and McKinnis, M. "Phonic Ability: Its Measurement and Relation to Reading Ability," School and Society, LI (1940), pp. 190-192.

# USING CHILDREN'S SCHOOL ATYPICALITIES TO INDICATE OCULAR DEFECTS \*

DOROTHY JONES KELLEY  
Chicago, Illinois\*\*

## Problem

THE PURPOSE of the study was to investigate the validity of using certain atypical school achievements, visual behaviors, and school behaviors for indicating the presence of ocular defects. Atypical, as used in this study, means "not showing characteristics of the group" in an acceptable or desirable way. Ocular defects refer to anomalous functioning of the refraction and fusion mechanisms of the eyes. Anomalies in visual perception or psychological maturation were not considered.

Educational research has shown that some children may achieve well and adjust well in school life in spite of visual defects. It also has indicated that the following may be symptomatic of ocular defects for individual children: School achievement that does not fully represent a child's true ability, visual behaviors including manipulation of reading materials that are anomalous, and school behavior that is not desirable.

To investigate this implication selected findings from a child observation project carried on at the Park Manor School, Chicago, over a period of a year and a half, were used as the research data for this study. The 553 children, 286 boys and 267 girls, were largely of Scandinavian and Irish descent and appeared to represent families of average socio-economic background. Their school grades ranged from grade 1A, the second half of the first grade, to grade 6B, the first half of the sixth grade. All of the children within these grades were used as subjects.

## Instruments

The research data were obtained from four sources: the administration of the Massachusetts Vision Test by the writer; cumulative school achievement records; the writer's observations of children's visual behaviors, that is, their physical adaptations to seeing; ratings of school behavior by fourteen classroom teachers.

\*The complete report of this study appears in the author's Doctor's Dissertation, "The Validity of Using Children's Atypical School Achievements and Behaviors for Indicating the Presence of Ocular Defects," Northwestern University, 1949.

Massachusetts Vision Test findings were sorted into a twofold classification of passing or failing. The 274 children passing the three sub-tests (Visual Acuity Test, Plus Sphere Test, Maddox Rod Tests) were classified as children with normal vision. The 279 children failing one or more of the sub-tests were classified as children with disabled vision.

Atypical school achievements 1, investigated as possible indicators of ocular defects, included (a) retarded rate of school promotion (one or more semester-grades repeated); (b) less than ten months reading achievement growth in one year<sup>2</sup>; (c) lowest quartile reading achievement placement in classroom group; (d) four months or more minus deviation from mental age grade expectancy in reading achievement; and (e) teachers' grades not indicating "Efficient" reading for semester prior to and for semester at time of vision testing.

Visual behaviors, as a term, describes the observable fusion adjustments of the children's eyes, their reading distance adjustments, and other observed physical factors. Atypical visual behaviors, investigated as possible indicators of functional eye defects, included reading distance measured as too-short or too-long<sup>3</sup>, atypical binocular following<sup>4</sup> (pursuit), and nearpoint convergence<sup>5</sup>, atypical peripheral eye appearance or complaint of pain or dysfunction<sup>6</sup>, tropia (crossed eye), tilted head posture, and book reading distance, observed as too-short or too-long for children seated at their own desks.

Atypical school behaviors, investigated as possible indicators of eye defects were those for which numerical ratings fell above the medians

1. School achievement data were taken for the semester in which vision test findings were procured and for the preceding semester or year. Ranges for data were, respectively, as lettered above: (a) Three semester-grades repeated to two semester-grades accelerated, (b) 0 months to 4.2 years growth in one year, (d) 2.7 years below grade expectancy to 3.5 years above, (e) two final grades of "Not Improving" to two final grades of "Efficient".
2. A child's expected rate of learning, which depends upon the relation between his intelligence quotient and his level of academic achievement, was disregarded since the coefficients of correlation for the covariants, reading achievement growth in one year and the vision test findings, did not vary significantly when the I.Q. as a factor was held constant. (Kuhlman-Anderson Tests)  $\phi = .16$  for 69-89 I.Q.'s ( $N = 27$ );  $\phi = .08$  for 90-109 I.Q.'s ( $N = .84$ );  $\phi = .04$  for 110-139 I.Q.'s ( $N = .35$ )
3. Measured reading distance for each child was determined away from his classroom while he read aloud from paper covered material in an optimum seating and lighting environment. This distance, perpendicular from the plane of the page to the outside corner of his eye, was measured with a ruler. His reading distance was considered normal if

in six ranges of computed behavior rating scores, averaged from two sets of teachers' ratings on the Haggerty-Olson-Wickham Rating Schedules, procured in successive semesters. These six ranges consisted of total scores for Schedule A (Behavior Problem Record) and Schedule B (Behavior Rating Scale) and of four sub-total scores related to intellectual, physical, social, and emotional traits.

### Method

The covariational method of correlation was used to investigate the validity of using the five atypical school achievements, seven atypical visual behaviors, and six atypical school behaviors, either as single indices or in series, for indicating the presence of ocular defects. Phi ( $\phi$ ) coefficients of correlation were computed to determine the degree of association between any two selected variables, one of which was always vision test findings (Table I).

For example, normal and retarded rate of school promotion findings were plotted against normal and disabled vision test findings, in a four cell graph. Similar phi coefficients were also computed for the children separated by grade into three groups, 1A-2A grades, 3B-4B grades, and 4A-6B grades (Table I, Column II).

Validity of the Massachusetts Vision Test, the validating criterion of this study, was indicated by a coefficient of contingency of .61, computed by the writer from data reported by Sloane (9). A coefficient of contingency of .76 indicated the reliability of the vision test in measuring the visual acuity of right eyes of 287 of the Park Manor School children on two administrations, a year's time intervening.

### Summary of Findings

The following findings imply answers to two questions formulated in the study:

it did not vary more than one and a half inches shorter or longer than the length of his forearm measured with a tape measure. (Reference 2, 212-213) Atypical distances ranged between 6 to 1½ inches shorter and 1½ to 6 inches longer than forearm lengths.

4. Binocular following ability was determined by holding a piece of white chalk about two feet from a child's eyes and by moving it from left to right slowly and then in a circle, while requiring him to follow the movement with both eyes. Cases ranged from those having no difficulty to those having no pursuit control at all. Any difficulty was considered atypical.
5. A child's nearpoint convergence was checked by holding a piece of chalk about eighteen inches from his eyes and by moving it slowly

TABLE I  
SELECTED INDICES RELATED TO CHILDREN'S VISION TEST FINDINGS  
(Listed in Order of O Correlation Values for all Grades)

Selected Indices	Grade Groups							
	All Grades		1A - 2A		3B - 4B		4A - 6B	
	N	Φ	N	Φ	N	Φ	N	Φ
<b>School Achievement Indices:</b>								
Rate of School Promotion	543	.10	209	.07	151	.12	183	.07
Reading Achievement Growth in One Year	346	.06	53 <sup>a</sup>	-.21	129	.10	164	.10
Reading Achievement Quartile Placement in Classroom Group	530	.05	200	-.07	149	.14	181	.12
Teachers' Reading Grades for Two Semesters (Averaged)	535	.04	203	.04	151	-.10	181	-.04
Reading Achievement Deviation from Mental Age Grade Expect- ancy	449	.03	126	.08	143	.02	180	.00
<b>Visual Behavior Indices:</b>								
Measured Reading Distance	553	.33	218	.33	152	.29	183	.36
Atypical Eye Appearance or Complaint	553	.21	218	.11	152	.14	183	.31
Presence of Tropia	553	.12	218	.15	152	.12	183	.10
Observed Reading Distance	343	.12	48 <sup>b</sup>	.19	134	.16	161	.06
Binocular Following	534	.10	215	.11	145	.15	174	.06
Head Posture	553	.06	218	.06	152	.08	183	.06
Binocular Convergence	553	.04	218	.04	152	.03	183	.09
<b>School Behavior Indices:c</b>								
Social Traits	503	.18	182	.11	149	.13	172	.04
Total Schedule A	503	.18	182	.26	149	.09	172	.03
Total Schedule B	503	.10	182	.05	149	.28	172	.01
Intellectual Traits	503	.07	182	.04	149	.20	172	.06
Emotional Traits	503	.04	182	.04	149	.18	172	-.06
Physical Traits	503	.04	182	.04	149	.18	172	-.05

<sup>a</sup> 2B - 2A grades only.

<sup>b</sup> 2A grade only.

<sup>c</sup> Two averaged teachers' ratings on the Haggerty-Olson-Wickham Behavior Rating Schedules,  
A and B, and four sub-divisions of Schedule B.

- A. How validly can atypical school achievements and behaviors be used for indicating the presence of ocular defects for school children?
1. Phi coefficients of correlation between vision test findings and findings of five school achievement indices ranged from .08 for the index, reading achievement deviation from mental age grade expectancy, to .10 for the index, rate of school promotion. Similar phi relationships for the seven visual behavior indices ranged from .04 for binocular convergence to .33 for measured reading distance. For the six school behavior indices phi values ranged from .04 for Social Traits to .18 for Physical Traits (Table I).
  2. The highest phi value regarding the relationship between the vision test findings and the findings of any single index occurred for the visual behavior index, measured reading distance. It was .33 (Table I).
  3. Phi coefficients of .21 and .12 occurred for relationships between vision test findings for all the children and findings for the index, atypical eye appearance and complaint, and the findings of the index, tropia. However, these two visual behavior indices screened correctly 96 percent and 87 percent, respectively, from the two small groups of children ( $N = 27$ ;  $N = 15$ ) who manifested these anomalies.
  4. The range was -.01 to .37 for phi correlation values between vision test findings and the findings by the last index in each of six selected series of school achievement indices, visual behavior indices, and school behavior indices<sup>7</sup> (Table II, Column I).

toward the bridge of his nose, while he was required to continue looking at it with both eyes. Atypical distances as determined in this study ranged from 4 to 18 inches.

6. Observed abnormalities in eye appearance were swollen or inflamed eyes, reddened or granulated eyelids, and eyes that appeared watery. Complaints were varied: My eyes hurt when I read too much; my eyes hurt when I look at the blackboard; sometimes this eye hurts at night; when I look at something on the blackboard it blurs; I close my eyes and then I can see it; my eyes are watery in the morning; when I read the letters get white and then black again; I couldn't see the blackboard in Miss \_\_\_\_\_'s room either; my eyes hurt when I read too close to my book; I told the teacher the letters go together and then go away.

5. Of the 33 children, selected by the first four indices in Series I, 20 had experienced retarded rates of school promotion. When their measured reading distance findings were correlated with vision test findings a phi coefficient of .40 resulted. In similar correlations, phi values of .24 and .55 resulted for 18 of the same 20 children when they were divided into 3B-4B and 4A-6B grades, respectively (Table II, Column I).
  6. Phi values of .26, .53, and .56 occurred as coefficients of correlation between measured reading distance findings and vision test findings for the three grade groups of children ( $N = 7$ ;  $N = 18$ ;  $N = 11$ ) selected as atypical by the last index in Series II, made up of school achievement indices (Table II, Column II).
  7. The highest phi coefficient, .80, occurred as the correlation value between measured reading distance findings and vision test findings for the group of eight 3B-4B children selected as atypical in the two visual behaviors, binocular convergence and binocular following, of Series IV (Table II, Column II).
- B. What are the chances that these indices would screen out children with ocular defects for referral or corrective care?
1. The chances that atypical school achievements, atypical visual behaviors other than measured reading distance, and atypical school behaviors would screen correctly children with ocular defects were, respectively, one in three, one in three, and even. (Percentages averaged for respective classifications of indices, above, were 33 percent, 33 percent, and 51 percent.) That they would select children with normal vision incorrectly for referral the chances were one in three, one in four, and two in five, 31 percent, 26 percent, 43 percent.
  2. The chances that reading distance measured as too-short or too-long would screen out children with eye defects correctly were
  7. For example, in Series I, children were first selected who were "Not Improving" in teachers' reading grades. These children were then screened for reading achievement deviation below mental age grade expectancy. Remaining children were further screened for first quartile classroom placement in reading achievement and then for less than ten months reading achievement growth in one year. Finally, for the 33 children remaining from this series of screenings, rate of school promotion findings—continuous versus retarded promotion—were correlated with their normal and disabled vision test findings, with a phi coefficient of .37 resulting.

TABLE II

FINDINGS OF LAST INDEX IN A SERIES RELATED TO CHILDREN'S  
VISION TEST FINDINGS

Number of Cases Screened and $\Phi$ Correlation Values	
Column I Series Selected	Column II Series Supplemented by Measured Reading Distance
<b>Series I:</b>  Low Teachers' Reading Grades Minus Reading Deviation M. S. Expectancy First Reading Achieve. Quartile Placement Atypical Reading Achieve. Growth, One Year Rate of School Promotion  All Grades: N = 33 $\Phi$ = .37	<b>Series I-A:</b>  Grades 1A-2A:              N = 2 Grades 3B-4B:              N = 11 $\Phi$ = .24 Grades 4A-6B:              N = 7 $\Phi$ = .55 All Grades:                 N = 20 $\Phi$ = .40
<b>Series II:</b>  Minus Reading Deviation M. A. Expectancy First Reading Achieve. Quartile Placement Reading Achievement Growth, One Year  All Grades: N = 91 $\Phi$ = .00	<b>Series II-A:</b>  Grades 2B-2A:              N = 7 $\Phi$ = .26 Grades 3B-4B:              N = 18 $\Phi$ = .53 Grades 4A-6B:              N = 11 $\Phi$ = .56 All Grades:                 N = 36 $\Phi$ = .48
<b>Series III:</b>  Minus Reading Deviation M. A. Expectancy Reading Achievement Quartile Placement  All Grades: N = 280 $\Phi$ = .11	<b>Series III-A:</b>  Grades 1A-2A:              N = 30 $\Phi$ = .35 Grades 3B-4B:              N = 43 $\Phi$ = .24 Grades 4A-6B:              N = 40 $\Phi$ = .30 All Grades:                 N = 113 $\Phi$ = .29
<b>Series IV:</b>  Atypical Binocular Convergence Binocular Following  All Grades: N = 120 $\Phi$ = .08	<b>Series IV-A:</b>  Grades 1A-2A:              N = 39 $\Phi$ = .48 Grades 3B-4B:              N = 8 $\Phi$ = .80 Grades 4A-6B:              N = 4 All Grades:                 N = 51 $\Phi$ = .51
<b>Series V:</b>  Atypical Social Traits Reading Achievement Quartile Placement  All Grades: N = 249 $\Phi$ = -.01	<b>Series V-A:</b>  Grades 1A-2A:              N = 46 $\Phi$ = .39 Grades 3B-4B:              N = 25 $\Phi$ = .41 Grades 4A-6B:              N = 26 $\Phi$ = .09 All Grades:                 N = 97 $\Phi$ = .30
<b>Series VI:</b>  High Behavior Problem Record Reading Achievement Quartile Placement  All Grades: N = 192 $\Phi$ = .03	<b>Series VI-A:</b>  Grades 1A-2A:              N = 35 $\Phi$ = .28 Grades 3B-4B:              N = 23 $\Phi$ = .07 Grades 4A-6B:              N = 19 $\Phi$ = .19 All Grades:                 N = 77 $\Phi$ = .24

two in three (62 percent) and those with normal vision incorrectly one in three (30 percent).

3. From the two groups of children ( $N = 20$ ;  $N = 36$ ) already selected as atypical by last indices in the Series I and II of school achievement indices (Table II, Column I) the chances were two in three (67 percent in both groups) that the visual behavior index, measured reading distance, would screen out correctly those with ocular defects and one in five (20 percent in both groups) that it would screen incorrectly those with normal vision.
4. From the 51 children selected, first, by atypical binocular convergence and then by atypical binocular following (Table II, Column I) the chances were four out of five (80 percent) that atypical measured reading distance would screen correctly those having ocular defects according to the vision test findings.

### Discussion

Examination of the phi coefficients of correlations in Tables I and II reveals usefully significant relationships between findings by the criterion for the study, the Massachusetts Vision Test, and findings for only one of the selected indices, namely, the visual behavior index, measured reading distance.

There may be general significance however, in the fact that all the relationships though low were consistently positive for the Park Manor School children as a single group (Table I). Also phi correlation values of significance did result when the indices were used in series and the children were grouped by grades (Table II).

Validity of School Achievement Indices for Screening Children's Vision.—There was no significant covariant relationships between the vision test findings and findings for any one of the school achievement indices. Though rate of school promotion appears first in the list of indices its phi value of .10 lacks statistical value (Table I).

Thus no one of the school achievement indices could be used as a truly valid indicator of the presence of eye defects for the Park Manor School children. This fact conforms with the findings of Monroe, Farris, Fendrich, and Kopel to the effect that many children achieve well in school in spite of vision difficulties (1, 4, 5, 6).

Validity of Visual Behavior Indices for Screening Children's Vision.—Slightly significant relationships between vision test findings and findings by one of the visual behavior indices, measured reading distance, were indicated by the phi coefficients of correlation of .33 for the 1A-2A children, .29 for 3B-4B children, and .33 for the children in a single

group (Table I). A percentage analysis also indicated this relationship thus: Atypical reading distance occurred for 68 percent of the children with ocular defects. Normal reading distance occurred for 70 percent of the children with normal vision.

The two indices, atypical eye appearance or complaint and the presence of tropia, had considerable significance for individual children, especially in grades 3B-4B, since 96 percent ( $N = 27$ ) and 87 percent ( $N = 15$ ), respectively, of the children who had manifested these anomalies, had ocular defects according to the vision test. But as devices for predicting the presence of eye defects they lacked validity because there were many more children who had eye defects who did not manifest either of these two anomalies.

It should be noted though that the phi coefficient of correlation for the index, atypical eye appearance or complaint, increased from .11 for 1A-2A children to .14 for 3B-4B children, and to .31 for 4A-6B children (Table I).

Validity of School Behavior Indices for Screening Children's Vision.—None of the findings by the six school behavior indices (averaged teachers' ratings on the Haggerty-Olson-Wickham Rating Schedules) showed any significant covariant relationship to the vision test findings. Thus there was no statistical evidence of validity for the use of any of these indices in screening the Park Manor School children's vision.

However, there might be a slight comparative significance in the fact that, with one exception, the highest phi values occurred for the 149 children in the 3B-4B grades (Table I).

Validity of Last Index in a Series for Screening Children's Vision.—Study of the phi values presented in Table II reveals some usefully significant relationships between the Massachusetts Vision Test results and findings from the use of indices in certain selected series when the visual behavior index, measured reading distance, was added as the last in a series and the children were grouped by grades.

For example, from the 33 children already screened atypically by the previous indices in Series I the school behavior index, rate of school promotion, screened correctly 75 percent of the children with ocular defects ( $N = 20$ ) and 61 percent of those with normal vision ( $N = 13$ ), the phi value being .37. When measured reading distance was added to make the Series I-A and the children were then grouped by grades, the phi coefficient of correlation increased from .37 (Table II, Column I) to .40 for all the children and to .55 for the seven children in the 4A-6B grades.

Similarly, when measured reading distance was added to make the Series II-A and the children were grouped by grades, the phi value increased from .00 to .26, .53, and .56 for the 2B-3A, 3B-4A, and 4A-6B grades respectively (Table II, Column II).

These increasing phi values for Series I-A and II-A imply an increase in the covariant relationship between measured reading distance

findings and the vision test findings for those Park Manor School children who had demonstrated atypical school achievements as their school grade placement advanced.

Phi coefficients of correlation for Series IV-A showed usefully significant relationships between the vision test findings and those of measured reading distance for the 51 children previously screened by the two visual behavior indices, atypical binocular convergence and atypical binocular following of Series IV. The phi values were .48 for the 39 children in 1A-2A grades and .80 for the eight children in 3B-4B grades (Table II, Column II). It is notable that all four of the 4A-6B pupils, similarly screened atypically in Series IV, had ocular defects according to the Massachusetts Vision Test. Measured reading distance screened correctly three of these four pupils.

There is indicated thus a significant relationship between measured reading distance findings and the vision test findings for the 12 Park Manor School children who still demonstrated atypical binocular convergence and following after leaving the 1A-2A grades.

In general, the findings presented for Series I-A, II-A, and IV-A (Table II), which indicate that more atypicalities in school achievement and visual behavior occur for children with ocular defects than for children with normal vision, may be compared with the research findings of Robinson for 30 seriously retarded readers in her study, Why Children Fail to Read. One of her conclusions was that the most seriously retarded readers showed the greatest number of anomalies—social, visual, and emotional—whereas the least retarded readers presented the smallest number of anomalies (8).

### Conclusions

1. In the absence of a more valid procedure, atypical measured reading distance may be used in an optimum environment as a simple measure for screening school children's vision. (See footnote 3.)
2. Atypical measured reading distance may be used with more useful validity for screening the vision of children who already manifest clusters of atypical school achievements and visual behaviors than for screening the vision of unselected children. Thus clusters of atypical school achievements and visual behaviors should be used as indicators of the need for testing the vision of the children who manifest them, especially beyond the primary grades.
3. In this study school achievement, behavior, and observable visual atypicalities, other than measured reading distance, were not significant as single measures in the screening of the children's vision. However, there is significance in the fact that the covariant relation-

ships though low were consistently positive for the Park Manor school children.

4. In view of the fact that the vision screening criterion of this study, the Massachusetts Vision Test, did not itself have high validity, the findings reported herein may be more significant educationally than they appear to be statistically.

#### BIBLIOGRAPHY

1. Farris, L. P. "Visual Defects as Factors Influencing Achievement in Reading," Journal of Experimental Education, V (1936), pp. 58-60.
2. Harmon, D. B. "Lighting and Child Development," Illuminating Engineering, XL (1945), pp. 199-232.
3. ———. "Some Preliminary Observations on the Developmental Problems of 160,000 Elementary School Children," Medical Woman's Journal, XXXIX (1942).
4. Fendrick, P. "Visual Characteristics of Poor Readers," Teachers College Contributions to Education, No. 656 (New York: Teachers College, Columbia University, 1935).
5. Kopel, D. Physiological Factors Associated with the Etiology of Reading Disability. Unpublished Doctor's Dissertation, Northwestern University, 1935.
6. Monroe, M. Children Who Cannot Read (Chicago: University of Chicago Press, 1932).
7. Notson, E. B. "Reading Distance and Some of Its Implications," School and Society, XXV (1927), pp. 353-354.
8. Robinson, H. M. Why Pupils Fail in Reading (Chicago: University of Chicago Press, 1946).
9. Sloane, A. E. "Massachusetts Vision Test, An Improved Method of Testing Eyes of School Children," Archives of Ophthalmology, XXIV (1940), pp. 924-939.

# IMPROVEMENT OF READING SKILLS THROUGH PREPARATION OF MATERIALS

RACHEL S. SUTTON  
University of Georgia  
Athens, Georgia

TEACHERS IN the elementary school have observed the value of pupil-prepared materials in learning to read. The beginning reader relates an incident, which the teacher records on the blackboard or reading chart, and he has the satisfaction of reading his story again and again. Oftentimes the slow and retarded reader makes progress when he writes his own basal text.

The investigator was concerned with the amount of progress a normal group of children might make under normal classroom conditions with an emphasis on writing stories, poems, and reports to be read later by other members of the class. Two groups were chosen in a semi-rural county. The experimental group was the fifth grade in the University of Georgia Elementary School, the control group was a fifth grade in the Winterville School, ten miles away. Children were transported to both schools on buses. The parents of both groups owned modest homes with gardens or small farms and worked in the industries of nearby Athens.

The test scores of the two groups used in the investigation are as follows:

	<u>Exper-</u> <u>imental</u>	<u>Control</u>
Average I. Q. (California Test of Mental Maturity)	89.6	86.7
Highest I. Q. (California Test of Mental Maturity)	115	114
Lowest I. Q. (California Test of Mental Maturity)	59	60
Percent of children of average intelligence	35.5%	36%
Percent of children above average intelligence	12.9%	4%
Percent of children below average intelligence	51.6%	60%
Average percentile rank in total mental factors (C. T. M. M.)	32.81	27.68
Average grade equivalent for all subjects (Metropolitan Test)	4.74	4.54
Average grade equivalent for reading (Metropolitan Test)	4.79	4.83
Average grade equivalent in reading capacity (Durrell-Sullivan)	4.75	3.79
Highest grade equivalent in reading capacity (Durrell-Sullivan)	7.7	5.6

Lowest grade equivalent in reading capacity (Durrell-Sullivan)	2.7	2.1
Range of grade equivalent in reading capacity (Durrell-Sullivan)	5.0	3.5
Median grade equivalent in reading capacity (Durrell-Sullivan)	4.4	4.0
Average grade equivalent in reading achievement in November (Durrell-Sullivan)	4.35	4.29
Highest grade equivalent in reading achievement in November (Durrell-Sullivan)	8.0	7.0
Lowest grade equivalent in reading achievement in November (Durrell-Sullivan)	2.2	2.5
Range of grade equivalent in reading achievement in November (Durrell-Sullivan)	5.8	4.5
Median grade equivalent in reading achievement in November (Durrell-Sullivan)	4.1	4.2

The children in the experimental group were encouraged to write about anything that interested them, to recount their own or others' actual experiences, or to use their imaginations in make-believe stories. A period each day was set aside for sharing with the class the stories written. After a story was read before the class, the teacher, in conference with the pupil, corrected it and discussed the boldest mistakes. The child then rewrote the story and incorporated the corrections and suggestions for improvement.

The second copy of the story frequently contained errors. A typist corrected all errors in the second draft of the story, so that the child could see in typed form a story free from errors in spelling, grammar, sentence, and paragraph structure. In this way the child was given the opportunity to grow into the correct usage of words, spelling, and grammar without being forced into it so fast that he became confused and discouraged. The typed stories were bound together in a loose-leaf notebook, which was always accessible to the class. Pupils were encouraged to read their own and others' typed stories whenever they had free time. This they did with apparent interest and pleasure.

The children were asked in the sharing period to evaluate each story as it was read. They stated at that time whether or not they liked the story read to them, but were unable to make many intelligent criticisms.

On the next day the teacher asked the question, "What must you have before you begin to write a story?" When some child answered, "An idea," she asked from what sources ideas were secured. She listed on the board sources of ideas as the class gave them to her. She asked where more information could be secured about ideas. The sources of information were also listed. The children brought out the necessity of

acquiring information. For example, a person might have an idea for an Indian story, but before he could write a good Indian story he would need to learn all he could about Indians—the way they dressed, lived, talked, acted, and the like. Great emphasis was placed on the value of giving details to make stories vivid and real, and on making the characters act and speak as do persons in real life.

The children learned that they not only must have ideas, and information about their ideas, but that they would need to develop them in a form interesting and intelligible to others. The teacher made a long list of the children's recommendations for improving a story. This list included such items as, proper sequence, use of words which are easily understood, sentences that are short and clear in meaning. In this connection, the children emphasized that every sentence should have a subject and predicate, and that it was better to have short sentences than to have long, rambling ones in which the subject became lost; and the word "and" was overworked.

After the three lists of suggestions, formulated by the class, were completed, the children were encouraged to use them in their own writing, and to appraise their own and others' stories by them. They were taken on trips to places on the University of Georgia campus, the telephone and telegraph offices, the creamery, the bakery, and to the fields and woods near the school in search of ideas and information. They were given more time in the library for reference work.

Whenever a story was read before the class, the students evaluated it. At first they conformed almost entirely to the lists they had formulated as a basis of judgment. Gradually the children became conscious of other values than those listed, and began to make use of them in analyzing the weak and strong points of a story. The teacher read to the class selections from children's classics and the pupils were encouraged to evaluate these also.

The class became sensitive to more discriminative vocabularies. The children made lists of overused words, and synonyms for them. They discovered how the use of a few "pet" words detracted from their writings, and how these overworked words could be replaced by other words not difficult and obtuse, but familiar and concrete.

Some of the best stories were selected and read before the entire school in an assembly period. As the children became more proficient in evaluating stories, they were encouraged to read one of the stories written during the first part of the experiment and then immediately afterwards to read one recently written. The author and the entire class together compared the two stories, and noted the ways in which improvements had been made.

During the whole of the experiment, no child was graded on his stories nor forced to write. The emphasis was placed on stimulation of interest, ways of making improvement, and on evaluation of the writings

by the children themselves.

The scores of both groups in the May testing are as follows:

	<u>Exper-</u> <u>imental</u>	<u>Control</u>
Average grade equivalent in reading achievement in May (Durrell-Sullivan)	5.25	4.64
Highest grade equivalent in reading achievement in May (Durrell-Sullivan)	8.8	7.5
Lowest grade equivalent in reading achievement in May (Durrell-Sullivan)	3.2	2.6
Range of grade equivalent in reading achievement in May (Durrell-Sullivan)	5.6	4.9
Median grade equivalent in reading achievement in May (Durrell-Sullivan)	4.9	4.3
Average grade equivalent of improvement during four months	.90	.35

Analysis of Test Results for the Experimental Group. — There were thirty-one children in the fifth grade of the University of Georgia Elementary School. This school, although it is administered jointly by the County and the University, is similar in most respects to other schools in the County. According to the California Test of Mental Maturity, 51.8 percent of those children were below average in intelligence, 35.5 percent were of average intelligence, and 12.9 percent were above average. (An intelligence quotient of 90-110 was considered average.) The lowest I. Q. in the group was 59 and the highest I. Q. was 115. The average I. Q. of this group of thirty-one was 89.6.

The group was given the Metropolitan Achievement Test on March 4, 1952. The average grade equivalent for all subjects was 4.74; the average grade equivalent in reading was 4.79.

According to the Durrell-Sullivan Reading Capacity Test, the average grade equivalent in reading capacity was 4.75. On November 20, 1951, Form A of the Durrell-Sullivan Reading Achievement Test was given; the average grade equivalent in reading achievement was 4.35. The reading improvement program was begun on February 18, 1952, and continued until May 13—at which time Form B of the Durrell-Sullivan Reading Achievement Test was given. The average grade equivalent in reading achievement on May 13 was found to be 5.25 grades. Therefore the average improvement was .9 grade. It was .5 grade beyond the capacity grade equivalent for that group. This progress had been made from November 20 to May 13. During that time, school was closed for three weeks, two weeks for Christmas and one week because the roads were impassable. The .9 grade improvement was made over a period

of four months, during three of which the reading improvement program was in effect.

No progress was made by two children; in fact, the grade equivalent was .3 grade lower for one child and .7 grade lower for the other. Both children had visual deficiencies and this apparent lack of progress may have been due to difficulty in reading the tests. Both children had entered into the program very willingly; and the boy who retrogressed .7 grade, had written thirteen stories with enthusiasm.

The most improvement was made by a girl with an I.Q. of 115; 2.2 grades. The one who stood second in the amount of progress made was a boy with an I.Q. of 59. His progress was 1.8 grades.

Analysis of Test Results for the Control Group. — A group of 25 children from the fifth grade of the Winterville School in Clarke County was tested, but no concentrated effort to improve reading was made in that school. According to the California Test of Mental Maturity, 60 percent of those children were below average in intelligence; 36 percent were of average intelligence, and 4 percent were above average. (An intelligence quotient of 90-110 was considered average.) The highest I.Q. in this group was 114 and the lowest was 60. The average I.Q. of this group of 25 was 86.7.

The group was given the Metropolitan Achievement Test on March 12 1952; the average grade equivalent for all subjects was 4.54 grades; the average grade equivalent in reading was 4.83.

According to the Durrell-Sullivan Reading Capacity Test given in November, 1951, the average grade equivalent in reading capacity was 3.79. Also in November, Form A of the Durrell-Sullivan Reading Achievement Test was given; the grade equivalent in reading achievement was 4.29.

On May 11, Form B of the Durrell-Sullivan Reading Achievement Test was given; the average grade equivalent in reading achievement was found to be 4.64. The average improvement of reading achievement in grade equivalent was found to be .35. This progress was made from November 21 to May 12, but during that time, school was closed for three weeks, two weeks for Christmas and one week because the roads were impassable. The improvement of .35 grade equivalent was made over a period of approximately four months.

No improvement at all was made by eight of the twenty-five students. Seven made less on the May test than on the one given in November. The greatest improvement by a single child was 2.1 grade equivalent, a girl with an I.Q. of 86. The child who ranked second in amount of improvement had an I.Q. of 81. The child with an I.Q. of 114, highest in the group, made .7 grade improvement.

Summary. — The progress made in the experimental group over a period of four months, according to the Durrell-Sullivan Achievement Tests, was .9 grade equivalent. The progress made by the control

group during the same period was .35 grade equivalent.

In the analysis of individual improvement, 6.45 percent of the experimental group and 32 percent of the control group made no progress.

# THE AMOUNT OF FREE READING ENGAGED IN BY INTERMEDIATE GRADE PUPILS WHO HAVE VIEWED TELEVISION FOR ONE YEAR OR MORE

CHARLES A. WELLS

TIMOTHY J. LYNCH

American International College  
Springfield, Massachusetts

THE LINE of investigation in this study was suggested by a comprehensive paper written by Dr. Paul Witty (1), "Television and the High School Student." In the course of his article, Professor Witty reports an unpublished study of high school students by Lewis which demonstrates that amount of television viewing is related inversely to length of television ownership. This thought-provoking bit of information, as a paper by Morgan (2) points out, is supported by student opinion. If the same situation holds good at the intermediate level, and there is no very good reason why it should not, what are these youngsters who are no longer viewing television so avidly doing in the way of reading? To answer this question we have chosen to investigate the free reading of a group of intermediate pupils who come from homes in which there has been television for one year or more.

## Review of the Literature

Witty (1) has already reviewed most of the studies which are significant in terms of our purposes. However, omitted by Witty in his survey of the literature, there has been reported the results of a survey of 5637 urban homes by Batten, Barton, Durstine, and Osborn, Inc., (3) which compares the amount of reading done in television owning homes with that done in non-television owning homes. Although the percentage reading favors the non-television homes, the overall percentage differences are not striking, and the findings contain seminal observation that much of the time spent in television viewing is time which was formerly devoted to radio.

## Sample and Procedures

Witty (1, p. 250) writes, "Yet it has been shown that children and youth are taking television in stride and are making adjustments which, in many cases, result in a successful assimilation of television in their total patterns of interest and activity."

And Hutchinson (4) says, "Those who would read, and have a purpose in doing so, will read well and extensively."

Hutchinson's statement boils down to the observation that those who prefer reading over any other activity will continue to do so. If one makes due allowance for the changes and vicissitudes of existence, this would seem by definition to be a secure position. Similarly, Witty's use of the qualifier "successful" still leaves one with the problem of defining that expression in terms of what has happened to the youngster's reading habits following this period of intensive television viewing.

Perhaps the best way in which to verify Hutchinson's judgment would be an individual longitudinal approach, since statements about individuals are involved. Similarly, Witty's "successful assimilation" would seem to demand a careful investigation of the personality gestalt of the youngsters, preferably of the case study type.

The exigencies of necessity, however, made it impossible to follow either of these approaches, however desirable they might be. Rather, the writers chose to accept evidence on the question which could be gathered and evaluated somewhat more expeditiously. Arbitrarily, it was decided that the aspect of reading habits which would be investigated would be time devoted to free reading or, in other words, reading done voluntarily and not as a matter of school assignments. The prerequisite for television ownership was that the set must have been in the home for a year or more. In order to eliminate intelligence as a factor only those subjects were used whose Kuhlmann-Anderson IQ's fell within the range 90 - 110. Our conclusions were based upon two types of evidence. First, there was our subjective judgment (based upon numerous studies of the amount of leisure time reading of students from first grade through college) of the adequacy of the amount of free reading reported by the student. Second, we looked for evidence of a negative correlation between the amount of time spent in television viewing and the amount of time devoted to free reading.

The group with which we began consisted of 158 intermediate grade children in the fifth and sixth grades of one school in a residential suburb of Baltimore. The socio-economic status of the suburb might be inferred from the fact that of the 134 who answered the question "Do you have a television set at home?", 113, or 81 percent, replied that they did. Since the elementary schools of this suburb are zoned, all of the children lived within a one mile radius of the school. Good reception from three television stations has been available in this area for over forty months.

The measuring instruments used were the Kuhlmann-Anderson Intelligence Test, a reading grade which was obtained by averaging the word meaning and paragraph meaning scores of the 1940 edition of the Stanford Achievement Tests, and a questionnaire. This latter was designed

to investigate the type of program enjoyed and the time devoted both to radio and to television, the time spent in free reading, sources of reading material, time devoted to comic books, and time spent in engaging in other recreational activities. The questionnaire also provided a form for evaluation of the type of material read. These rather specific questions provided a cross checking device for estimating the validity of student statements.

As happens so frequently, many of the questionnaires were of dubious validity and had to be excluded, some of the original group left the neighborhood while the study was in process, and the scores for some of the others were incomplete. Since, in addition, we were *a priori* restricted to students who could satisfy the requirement for length of television ownership, the sample upon whom the study was finally based consisted of 52 of the original number of 158. Both grades were considered together as if they constituted a single group.

### Results

The reading averages from the Stanford Achievement Tests ranged in grade level from 3.9 to 9.7. Range of television observation in hours weekly was from 3 - 42 with a mode falling between 9 and 11 hours weekly and a mean of 12.8. This is in line with the observations of Lewis and Morgan. Weekly hours of free reading ranged from 2 - 24 with a mode lying between 4 and 5 hours. In view of the many studies in this area which have obtained results no higher than this, this seems to be a very respectable figure and indicates that these youngsters spend at least a fair amount of time in free reading. Perhaps this represents a recovery from the initial impact of television. Perhaps, as suggested by Hutchinson, it represents a personal predilection for reading. Perhaps, however, considering the fallibility of human reports, some doubt should be attached to the statements of those who claim that they read less since the advent of television. It might be, as the survey by Batten, Barton, Durstine, and Osborn, Inc., suggests, that much of the time devoted to television viewing is time taken not from reading, or from reading alone, but from other activities as well.

Zero order  $r$ 's were: (1) between hours of television viewing and hours of free reading,  $r = .14$ ,  $\sigma_r = .13$ ; (2) between hours of television viewing and the reading average,  $r = .25$ ,  $\sigma_r = .13$ ; (3) between hours of free reading and the reading average,  $r = .28$ ,  $\sigma_r = .13$ . This latter  $r$ , while significant, is hardly so impressive as one might expect considering the nature of the functions involved. It is problematical, however, as to whether or not this is as high a coefficient as one might reasonably expect when one applies simple correlational techniques to a multivariate process such as reading. In any case, we did not obtain

the negative correlation which we judged to be necessary if we were to have evidence of the lasting impact of television viewing upon free reading.

To investigate some of the possible interrelationships among these variables, partial r's were computed. The partial r between hours of television viewing and hours of free reading, holding reading average constant, was .16,  $\sigma_r = .14$ . For hours of television viewing and the reading average, holding constant hours of free reading, the partial r was .23,  $\sigma_r = .12$ . As is immediately obvious, neither of these partial r's is acceptable at the 5 percent level of confidence, nor do they differ from the Product Moment r's for the same variables.

To dramatize the non-significant zero order coefficient between weekly hours of television viewing and weekly hours of free reading, r biserial was computed between these two variables, dichotomizing hours of television viewing at twelve hours. As was to be expected,  $r_{bis}$  equalled .02.

### Summary and Conclusions

Because there are studies indicating that with increased length of television ownership, time of viewing decreases, we set ourselves the task of investigating the amount of free reading done by intermediate grade children in homes in which there has been television for one year or more in order to determine what the effect is upon reading of long-time television ownership. As evidence, we used the statements of the children themselves. Also, we felt that if television has had lasting deleterious effects upon reading, there should be a negative correlation between hours of television viewing and hours of free reading. From the answers of the children to our questionnaire, we may conclude that after a minimum of one year's television ownership, our sample spends an average, although perhaps not a desirable, amount of time in free reading. Nor were negative correlations found between the variables to indicate continued harmful effects to reading from television viewing. To this extent, then, we offer empirical evidence to substantiate the statements cited earlier from Hutchinson and Witty. We trust that it will be of assistance to these men who are attempting the dual task of, on the one hand, answering the critics of television, and, on the other hand, of promoting the constructive possibilities of the medium.

## REFERENCES

1. P. Witty, "Television and the High School Student," Education, LXII (December 1951), pp. 242-251.
2. D. Morgan, "Television versus Reading," Wilson Library Bulletin, XXVI (December 1951), p. 327.
3. What's Happening to Leisure Time in Television Homes? (Chicago: Batten, Barton, Durstine and Osborn, Inc., 1951).
4. J. R. Hutchinson, "Reading, Writing and Television," Secondary Education, XIV (April-June 1949), pp. 9-10.

## Research Abstracts and Bibliographies

T. A. Lamke, Iowa State Teachers College, Cedar Falls, Iowa

A Further Analysis of a Follow-Up Study of 1820 June, 1948, Graduates from Nine Los Angeles City High Schools. Publication No. SC-453. (Los Angeles: Curriculum Division, City School Districts, 1951), 55 pp., n. p.

Reports findings from a survey made from six to ten months after high school graduation. Discusses implications for curriculum content.

Anderson, Walter A. and others. The Workshop Handbook (New York: Bureau of Publications, Teachers College, Columbia University, 1953), 65 pp. n. p.

Contains practical suggestions for planning and conducting workshops in education. Includes an annotated bibliography of 23 items.

An Effective Education Program in Kentucky. Bulletin of the Bureau of School Service, Vol. 25, No. 3. (Lexington: University of Kentucky, 1953), 112 pp., 50¢.

Reports proceedings of the twenty-ninth annual education conference and the eighteenth annual meeting of the Kentucky Association of Colleges, Secondary, and Elementary Schools.

Bowman, Paul H. and others. Studying Children and Training Counselors in a Community Program. Supplementary Education Monographs No. 78, June, 1953. (Chicago: University of Chicago Press, 1953), 136 pp. \$1.50.

The second of a series of reports on the work of a community youth development program designed to extend over several years. Recounts methods used to discover gifted and emotionally disturbed children; describes training of counselors and their activities.

Bridging the Gap Between School and College. (New York: Fund for the Advancement of Education, 575 Madison Ave., 1953), 127 pp., n. p.

A progress report on four related projects supported by the Fund for the Advancement of Education: (1) The efforts of the preparatory schools of Andover, Exeter and Lawrenceville, and Harvard, Princeton and Yale Universities, to make the last two years of preparatory school and the first two years of college a continuous process; (2) The collaboration of the public school system of Portland, Oregon, and the faculty members of Reed College on a city-wide program for enriching educational opportunities for public school children of exceptional ability; (3) Kenyon College's program of admission of high school graduates to advanced standing; (4) The early admission to twelve colleges and universities of selected students before high school graduation.

Kinker, H. Robert and William H. Fox. A Study of High School Guidance Services in a Six-State Area. Bulletin of the School of Education, Indiana University, Vol. XXVIII, No. 6, Nov., 1952. (Bloomington: Indiana University Bookstore, 1952), 108 pp., \$1.00

A sampling study of high school guidance services available in Illinois, Indiana, Kentucky, Michigan, Ohio, and Wisconsin.

Miller, Norman C., and David Lilien. An Early Experiment in Adult Education. (New Brunswick, N. J.: Rutgers University Press, 1953), 40 pp., A brief history of the Rutgers University Department of Extension from 1891 to 1903.

Proceedings, 1952 Invitational Conference on Testing Problems. (Princeton, N. J.: Educational Testing Service, 1953), 138 pp., n. p.

Records papers and discussions of the 1952 conference in New York City.

Romualdez, Sister Bellarmine. The Concept of Being in Modern Educational Theories. (Washington, D. C.: Catholic University of America Press, 1952), 208 pp., \$2.25

A study of the various positions taken by representative contemporary educational leaders regarding this fundamental philosophical concept. Includes an extensive bibliography.

Studies in Education 1952. School of Education Indiana University Thesis Abstract Series No. 4, 1953. (Bloomington: Indiana University Bookstore, 1953), 325 pp., \$1.00

Abstracts of the doctoral theses of all individuals who received the D.Ed. degree, and of some who received the Ph.D. in education, at Indiana University, 1952.

Sullivan, Sister M. Saint Catherine. The Concept of Authority in Contemporary Educational Theory. (Washington, D. C.: Catholic University of America Press, 1952), 239 pp., \$2.75.

Considers the backgrounds of the concept of authority in educational theory, and the relation of the concept to education in American democracy. Includes an extensive bibliography.

Travers, Robert M. W., and others. Exploratory Studies in Teacher Personality. Division of Teacher Education Office of Research and Evaluation Publication No. 14. (New York: College of the City of New York, 500 Park Ave., n. d.), 56 pp., n. p.

Reports 7 exploratory studies attempting to relate teacher personality to teacher effectiveness.

Traxler, Arthur E., editor. Modern Educational Problems. (Washington, D. C.: American Council on Education, 1953), 147 pp., n. p.

A report of the 17th educational conference held under the auspices of the Educational Records Bureau and the American Council on Education, at New York City. Contains papers and reports of panel discussions.

# JOURNAL *of EDUCATIONAL RESEARCH*

---

Volume XLVII

March, 1954

Number 7

---

## THE MECHANICS OF QUESTIONNAIRE CONSTRUCTION

JOHN E. NIXON  
Stanford University

THE PURPOSE of this article is to present concrete suggestions concerning the mechanical make-up and physical arrangement of questionnaire and check-list forms. This discussion is not concerned with the justification, purposes, strengths, weaknesses and other aspects of the questionnaire as a research instrument which are treated very thoroughly by a number of authors.

The present article is limited to a series of practical suggestions to be considered by the researcher in the actual construction, development of the physical form, and final publication of the questionnaire prior to submission to respondents.

### Importance of Form and Appearance

The suggestions which follow all relate directly to assisting the investigator in obtaining the largest possible percentage of replies to his check-list or questionnaire. It is generally agreed by the authorities in educational research that more returns will be forthcoming when questionnaire forms are attractively presented and easy to read and mark.

It is the conviction of the author that inclusion in a form of one or more of the suggestions listed herein probably will elicit a greater number of useable responses than would be obtained by ignoring the suggestions or by carrying them out only in part.

As Troops so clearly demonstrates, the primary objective of the investigator should be to achieve 100% response.<sup>1</sup> The suggestions below should assist the investigator in realizing that objective.

1. Walter S. Monroe, editor, Encyclopedia of Educational Research, (New York: The Macmillan Company, 1950), pp. 948-951.

Paper and Ink

1. The questionnaire should be placed on high quality paper.
2. The page size should conform to that which will be used in writing up the study.
3. Page size should be selected for ease in folding and mailing in standard number 7 and 8 envelopes.
4. At the discretion of the investigator, in accordance with the purpose of the questionnaire, and with deference to respondents, the use of colored paper (e. g. light green, yellow) sometimes is attractive and desirable.
5. The use of blue ink or other colored ink may enhance the appearance of the form.
6. The form should be printed if feasible. The attractiveness of a printed form over a mimeographed or typed form cannot be over-emphasized. If the investigator is seriously concerned with obtaining a high percentage of returns, printing is recommended. Printing has the additional advantage of using various sizes and types of print to attract attention by setting off directions, questionnaire items, and other special features on the form for ease of reading and marking by the respondent.

Arrangement

1. The first or covering page of the questionnaire should contain the complete designation of the sponsoring agency centered at the top. If at all possible questionnaire investigations should receive the sponsorship of some recognized professional organization rather than to be submitted only in the name of some individual person.
2. The date of mailing should appear on page one.
3. The consecutive numbering of each questionnaire is recommended. The number assigned should appear on every page of the questionnaire and in the same location, preferably near the upper right hand corner.
  - a. Use of numbered questionnaires gives the investigator an accurate record of all questionnaires returned in the event a respondent fails to write in his name or institution.
  - b. Another advantage of numbered questionnaires is in tabulation. Instead of using tally marks, the code numbers of respondents can be placed after each item as tabulated. These numbers can then be counted. This system requires more time for tabulation but gives the investigator a consolidated report of replies to each item by respondent without having to go back through each questionnaire to find how specific individuals answered a particular item. Further, this procedure expedites the writing of the

report when it is desired to name individuals or institutions answering given items in certain ways.

- c. Use of numbered questionnaires cannot be construed as under-handed or unethical by respondents if the investigator also includes a space for the writing in of the name of the respondent and the institution he represents.
- d. It is also suggested that each number be placed on two questionnaires for follow-up use.
4. A different questionnaire number should be assigned to each respondent included in the study, the numbered questionnaires being mailed to respondents according to the number assigned.
5. Provide a space for the respondent to fill in as he completes the form.
6. Provide space for the respondent's name and title.
7. Provide space for the organization and location of the respondent.
8. Include a statement giving the respondent an opportunity to complete your questionnaire with the assurance of confidential treatment of his replies if he so desires. For example, "May your name and that of your institution be mentioned specifically in connection with the data you have submitted? YES        NO       " Then, by all means, respect the desires of the respondent. It is the writer's experience that this technique will insure some confidential returns which can be used in the over-all report of data, which might otherwise never have been submitted.
9. Another technique for increasing the number of returns is to offer a summary of the findings of the study. For example, this line might be included, "Do you desire a summary of the findings of this study? YES        NO       " Obviously, you are obliged to send such a summary to those respondents who check this line in the "YES" space. A majority of respondents probably will check "YES."
10. It is frequently helpful to ask respondents to enclose supplementary charts, diagrams, figures or other materials not specifically asked for in the items themselves. However, experience shows that if such materials and data are essential to the study they should be obtained from respondents as answers to specific items within the questionnaire proper rather than to be requested as additional material to be enclosed.
11. The title of the study should show prominently near the top of the first or covering page.
12. It is recommended that the words "questionnaire" or "check-list" not appear on the form at all. It is well known that the questionnaire and check-list are used to such a great extent that many respondents do not bother to answer them any more. Reference to the questionnaire as a "form" or "instrument" may have a less irritating effect on certain respondents.

13. If the form can be accomplished in a relatively short time, for example, in fifteen to thirty minutes, it is recommended that this information be placed in a prominent location on the first page near the top. Respondents may be induced to answer the form if they believe it will not require too much time to do so. It is also suggested that this fact be emphasized early in any covering letter which may accompany the form before they throw the form in the wastebasket.
14. Include complete information about the recorder or compiler to whom the form should be returned, including full name, title, and complete address. This information should appear on the first page of the form if possible, otherwise, place it as the last item on the form. Do this even if a self-addressed envelope is enclosed for the return of the form because frequently this envelope is lost before the respondent answers the form.
15. The first page should contain a request for the return of the completed form (1) "at your earliest convenience," or (2) "on or before January 15, 1950." The first suggestion is favored on the assumption that the investigator has allowed sufficient time for the return and tabulation of data without putting pressure on respondents which they resent. However, many investigators prefer to place an actual, reasonable date prior to which they hope the respondent will return the form. The reasoning is that such a definite commitment date will encourage promptness on the part of respondents. The latter suggestion also is desirable when the investigator definitely is faced with a time limit for the tabulation, analysis, and reporting of the data.
16. On the first or covering page it usually is desirable to mention the inclosure of a self-addressed stamped envelope for returning the form.

#### Directions

1. Keep the directions as brief as possible. Try to construct items in such a manner that the method for marking them is self-evident.
2. Do not include in the same section of the form items requiring different types of response or marking. Set up the form in separate sections one for each type of written response required.
3. Provide for the mere checking of a possible answer already present on the form, where possible, rather than requiring a written answer.
4. Begin each new section of the form with brief instructions if necessary, then include a sample item with a sample response, if required for clarity.
5. Set off directions with a heavy, larger, distinctive lettering.

6. Space for Yes and No replies should be arranged vertically YES \_\_\_\_\_ rather than horizontally YES \_\_\_\_\_ NO \_\_\_\_\_.
7. Leave some space for comment after each major item, and encourage such comment if it will help to make the meaning of the reply more explicit. Do not expect to be able to tabulate information contained in comments, however.
8. Number every page of the form consecutively.
9. On the bottom of the next to last page insert the word "over" if there are more items for completion on the last or back page. Respondents sometimes will fail to look on the back page unless reminded to do so.
10. If several items are listed on successive lines, and each requires a check of some kind, make it easier for the respondent by skipping a line after every third item.

#### Covering Letter

1. Print the covering letter if possible. Otherwise, use the most attractive means of publication available.
2. Address the letter personally and specifically to individual respondents.
3. Arouse interest in the study briefly and quickly in the letter.
4. Use the letter head of the sponsoring agency, or of the sending institution.
5. Obtain the recommendation, sponsorship, or approval of noted authorities for inclusion in the letter, and their signatures at the bottom of the page along with that of the compiler.
6. Keep the letter short and on one page.

#### Envelopes

1. Have the return address of the compiler printed on the outgoing envelopes.
2. Print the address of the compiler on the return, self-addressed envelope.
3. Address envelopes in the same manner as the covering letter, to respondents personally by name.
4. The return envelope should be one standard size smaller than the sending envelope so that one will fit into the other without folding.
5. The questionnaire should be folded so that it may be returned in the smaller envelope without having to be refolded differently.

#### Mailing

1. By accurate use of letter weighing scales it can be determined just

what weight paper, and what weight and size envelopes can accommodate the questionnaire for minimum postage rates. For example, the writer was able to send two copies of a four page questionnaire, a covering letter, and a return envelope for 3 cents postage by weighing out the envelopes carefully, then determining just what weight paper was necessary to use for printing the questionnaire and the covering letter to stay within the 3 cent limit. Attention to such details easily can reduce postage costs fifty per cent or more.

2. It is recommended that two copies of the questionnaire be sent to each potential respondent, plus a covering letter and a self-addressed, stamped envelope. The covering letter should state that one copy of the form is for use as a work sheet and for file, and the other for return to the compiler. Many respondents desire to fill out one form roughly with pencil, then turn it over to a secretary for completion of the second form on the typewriter.
3. Most outgoing questionnaires, covering letters and self-addressed envelopes, can be accommodated in a U. S. Government No. 8 stamped envelope, unless the form is unusually large or lengthy.
4. A No. 7 U. S. Government stamped envelope should be used for return of the form to the compiler. The No. 7 envelope will fit into the No. 8 envelope without folding, and both No. 7 and No. 8 will hold regular 8 1/2 x 11 inch size paper folded two times lengthwise.

### Follow-up Letter

1. The follow-up letter likewise should be printed, or otherwise reproduced in the most attractive style possible.
2. The follow-up letter should contain one additional copy of the questionnaire and another self-addressed stamped envelope.
3. Do not attempt to send questionnaires for two cents by leaving the envelope flap open rather than sealed. If the study is worth doing, it is worth three cents in postage to assure proper care for valuable data.
4. It is recommended that follow-up letters be dispatched two weeks after the probable date of receipt of the original questionnaire by respondents. This procedure seems to be superior to waiting a month or longer before initiating follow-up requests.
5. Some authorities recommend sending as many as eight follow-ups in an effort to achieve 100% response.

### Summary

In summary, it will be noted in this list of suggestions for the mechanical construction of questionnaires that the emphasis has been on consideration for the respondent. The ultimate objective is to obtain

as many responses as possible, in the form of completed questionnaires which provide useable data. If questionnaire forms meet criteria of physical attractiveness and obvious consideration for the respondent, it is believed that the percentage of replies will be sufficiently high to fulfill the requirements of the investigator. Every conceivable inducement should be used in the hope of convincing one more potential respondent to take the time and effort necessary to answer the questionnaire.

# ART IN WRITING FOR EDUCATIONAL PERIODICALS: THE MAIN BODY

J. R. SHANNON \*  
San Diego State College  
San Diego, California

"THE HIGH lights of a discourse are the beginning and the end," said a former professor of English and reputable college president.<sup>1</sup> A top-ranking educational editor seconds the motion and stresses art in the first sentence and first paragraph as number one among his fifteen "tips on writing when you want to be read."<sup>2</sup> But emphasis on art in style of writing at these two places need not imply that the product of the means should be less than the product of the extremes. The big apples should not all be on top in the basket, or all on top and in the bottom; quality should be continuous, if not uniform, throughout.

## Background studies and present surveys

Earlier articles in this series dealt with art in writing the introduction and the ending,<sup>3</sup> and showed, incidentally, that more big apples were on top in the baskets than at the bottom. The present article concentrates on the main body, and reports two surveys, one of which analyzed the elements of excellence and elements of inexccellence in the main body of 193 magazine articles, and the other of which worked more extensively with 15 articles.

The first of the two earlier reports, the one on introductions, used 100 articles in each of 17 magazines chosen by graduate students on the basis of their majors or other interests. The second, the one on endings, used 100 articles in each of 16 periodicals similarly chosen. In the two groups of articles, one of 1700 and one of 1600, there were by sheer accident 193 articles common to both groups. These 193

\* The writer acknowledges assistance by several graduate students in serving as judges of expositions treated in this survey.

1. William Trufant Foster, Argumentation and Debating (New York: Houghton Mifflin Co., 1945), p. 221.

2. Rolfe Lanier Hunt, "When You Want to Be Read," Phi Delta Kappan, 29:44-45, September, 1947.

3. J. R. Shannon, "Art in Writing for Educational Periodicals: the Introduction," Journal of Educational Research, 44:599-610, April, 1951; and "Art in Writing for Educational Periodicals: The Ending," Journal of Educational Research, 46:333-345, January, 1953.

were studied more intensively for the present report on art in writing exemplified in the main body. The other 15 articles covered in the present report were all in a single issue of one periodical (Bulletin of the American Association of University Professors, Volume 36, Number 3, Fall, 1950).

In each of the two surveys reported herewith, the primary purposes were the same: (1) to identify elements of excellence which characterized superior style; (2) to identify elements of lack of excellence which characterized poor style; (3) to get a measure of graduates' students' competence in appraising quality and in discerning elements of technique which characterized superior style and poor style. An additional purpose in the analysis of the main bodies of 193 articles in the first survey reported herewith was to compare quality of style in the main body of the several expositions with that in the introductions and endings, as previously evaluated by other judges but not known by the students who evaluated the main bodies.

Definitions of levels of quality in both surveys were as follows:

Superior style is the unique characteristics of an author's method of communicating ideas in written form with an artistic quality which attracts admiration by itself, aside from the substance of the idea being expressed.

Good style is one in which ideas are expressed clearly and in a readable manner.

Poor style is one in which ideas are expressed in a way that is hard to comprehend and which tends to repel or annoy the reader.

Decisions in all cases were based solely on the quality of writing and not on the nature of the subjects discussed.

#### Students' competence at evaluating merit

Man will never know how the All Seeing Eye evaluates the style of writing in professional educational literature. Man does not know at this time how a jury of expert English critics evaluates it. But nevertheless, every reader evaluates. He evaluates either formally or informally, consciously or unconsciously, critically or uncritically. Such is inevitable. The judges who did the evaluating of style in the main bodies of the 208 expositions for the present report were mature college graduates, most of whom held responsible positions in public education. Their median age was above 35. They were carefully instructed and duly supervised. If they could not recognize merit when they saw it, and identify elements of finesse which constitute merit, it is probable that future efforts with similar juries will be fruitless.

There is considerable evidence indicating that the judges' evaluations were dependable -- not with equal and unanimous reliability,

but dependable on the whole. A statistical evidence is the summary of 50 judges' estimates of the 15 articles in the AAUP Bulletin. Forty-nine mature graduate students in education, plus the writer, their instructor in Research in Education, worked independently in deciding which three of the 15 articles were best written, and in analyzing and tabulating factors which made the top articles tops and those which kept the other articles from being tops.

A popular slogan in commercial advertising is: "Such a large number of people cannot be wrong." The probable mathematical validity of the slogan in its present application is attested by the trend of evaluations by the 50 judges. With 15 articles and only three to be chosen as tops, the law of averages would suggest that each article would get ten votes. But one article got 30 votes and another got 28; only three other articles got more than ten votes each.

If 50 archers shoot at the same target and their arrows cluster about the same point, it is unlikely that the place of convergence will be some eccentric point. The law of averages suggests that the archers are striking close to the bull's-eye.

Several evidences of reliability of judges' evidences are found in the data on the 193 articles common to the two earlier surveys. The main body of each of the 193 articles was evaluated independently by five separate judges. With five separate judges and three degrees of merit to distinguish (superior, good, poor), there were 21 possible combinations of judges' evaluations of the style in the main body of an article. The 21 possibilities are shown in the left vertical column of Table I, and to the right of each possible combination of evaluations is a figure showing how many times that combination occurred among the 193 articles.

The first evidence of the reliability of the judges' evaluations of the style of writing in the main bodies of 193 expositions is the absence of much radical disagreement among the judges, as shown in Table I. It is gratifying to find no instances at all of combinations 4 - 0 - 1, 2 - 0 - 3, or 1 - 0 - 4, and only three instances of other "sway-backed combinations." With variability of human judgments being what it is, as evidenced by Supreme Court decisions or by teachers' evaluations of pupils' school achievement, the 20 instances of complete unanimity of agreement among five judges, when they had three options per article, and the 113 instances of disagreement between only two adjoining levels of merit, are almost to be marveled at instead of deplored.

Further evidence of the reasonable reliability of the judges' evaluation of merit in the 193 magazine articles was obtained by personal verification of 38 extreme cases. The 38 fell into five groups: (1) articles which the judges had agreed unanimously in the earlier surveys had both attractive introductions and effective endings; (2) ones which the earlier judges had agreed unanimously had attractive introductions and ineffective endings, or unattractive introductions and effective

TABLE I

## FREQUENCIES OF COMBINATIONS OF JUDGES' EVALUATIONS OF MAIN BODIES OF 193 MAGAZINE ARTICLES

Combinations of Evaluations			Frequencies
S	G	P	
5	0	0	2
4	1	0	1
4	0	1	0
3	2	0	3
3	1	1	5
2	3	0	12
3	0	2	1
2	2	1	14
1	4	0	21
2	1	2	2
1	3	1	20
0	5	0	16
2	0	3	0
1	2	2	12
0	4	1	28
1	1	3	6
0	3	2	32
1	0	4	0
0	2	3	11
0	1	4	5
0	0	5	2
<b>Total</b>			<b>193</b>
<u>Summary</u>			
Unanimity of agreement			20
Two adjoining levels of merit			113
Three levels but with only one "erratic" evaluation			57
Sway-backed combinations			3
<b>Total</b>			<b>193</b>

endings; (3) ones which the judges of quality in the main body had unanimously rated superior, or unanimously rated poor; (4) ones on which the judges of main body had disagreed extremely (the three "sway-backed combinations"); (5) ones which the earlier judges had rated high on introduction and ending but the judges of main body had rated lower. The writer, who is exceptionally sensitive to style in writing, found himself in agreement with the majority of the five judges in 33 of the 38 extreme cases, in agreement with two of the five in two cases, in agreement with one of the five in one case, and in agreement with none of the five in two cases.

Common observation attests that tastes vary. Jack Sprat and his wife are not oddities. In at least five instances in which two judges disagreed, one calling a piece of writing superior and another calling it poor, the two cited the same characteristic as evidence in point. One of the five instances was the use of outline form for presenting details; another was use of quotations; a third was use of italics for emphasis; the fourth was use of first person; and the last was light and carefree style suitable for oral discourse.

"Judge not, that ye be not judged," is not wise counsel. Judging is inevitable, and through his judging the rater rates himself, willy-nilly. "Judge thoughtfully and analytically," is wiser counsel.

There was a slight experimental factor injected into the survey which threw some light on reliability of judges' evaluations. One graduate class was divided into teams of three each, each team judging the same articles among the 193, and another graduate class was divided into teams of two each, each team judging the same articles. The same instructions were given in each class except for one difference: the teams of two each received the same instructions as the teams of three each, but in addition were given a mimeographed specimen of what the writer regarded as superior writing and one of what he regarded as poor writing, and each specimen also had indicated the elements of each which made it superior or poor, as the writer saw them. Thus, the teams of two each were somewhat more thoroughly instructed and should have been somewhat more discriminating in their evaluations and analyses.

A comparison of the evaluations by the teams of two with those by the teams of three shows that:

1. Both members of the two-man teams agreed with two or three members of the three-man teams in 105 of the 193 articles.
2. Both members of the small teams agreed with one member of the larger teams in 19 cases.
3. Both members of the smaller teams classified articles as good, when the members of the larger teams had voted one superior, one good, and one poor, in 7 cases.

4. One member of the smaller team agreed with two or three members of the larger team, but the second agreed with none, in 40 cases.
5. One member of the smaller team agreed with one member of the larger, but the second agreed with nobody, in 8 cases.
6. Neither member of the small team agreed with anybody on the larger team in 14 cases.
7. The smaller (better instructed) teams were more discriminating in their evaluations than were the larger teams.
  - a. The ratio of number of elements of excellence cited to the number of articles said to be superior by the two-man teams was 3.60, whereas that by the three-man team was only 2.76.
  - b. The ratio of number of elements of inexcellence cited to the number of articles said to be poor by the two-man teams was 2.58, and that by the three-man teams was 2.39.
  - c. The smaller teams and the larger teams were approximately equally liberal in their evaluations. Counting a rating of superior as +1, a rating of good as 0, and a rating of poor as -1, a numerical value for each of the 193 main bodies was obtained. This score was raised by the smaller teams above what the larger teams had decided for 75 articles, lowered for 81, and left the same for 37.

#### Comparison of main bodies with introduction and endings

The earlier surveys (cited in footnote number 3) classified the introductions to the 193 articles as either attractive or not attractive and the endings as either effective or not effective. With three judges deciding on introductions and three others deciding on endings, there were 16 possible combinations of judges' evaluations of introductions and endings. These are shown in the left vertical column of Table II (3A and 3E means that all three judges classified an article's introduction as attractive and all three classified its ending as effective. N means either ~~not~~ attractive introduction or not effective ending.)

To the right of the 16 combinations in Table II there are shown in succeeding vertical columns: the frequencies of the 16 combinations; the mean point values for the articles in each bracket on quality of main body, as described under "7c" above; and the range of point values for the articles in each bracket.

Before analyzing Table II for comparisons of quality of writing in main bodies with that in introductions and endings of articles, some observations are in order on the relationship between quality of introductions and quality of endings:

TABLE II  
COMPARISONS OF MAIN BODIES WITH INTRODUCTIONS  
AND ENDINGS

Combinations of Evaluations of Introductions and Endings	F	Numerical Evaluations of Main Bodies	
		Mean	Range
3A and 3E . . . . .	1	1.00	
3A and 2E . . . . .	14	-0.21	2 to -4
3A and 2N . . . . .	17	-0.23	5 to -3
3A and 3N . . . . .	4	0.25	1 to -2
2A and 3E . . . . .	9	1.33	5 to -2
2A and 2E . . . . .	18	0.50	3 to -3
2A and 2N . . . . .	24	-0.46	4 to -4
2A and 3N . . . . .	15	-0.93	2 to -5
2N and 3E . . . . .	5	-0.40	3 to -2
2N and 2E . . . . .	10	-0.10	2 to -4
2N and 2N . . . . .	22	-0.77	2 to -5
2N and 3N . . . . .	14	-1.14	1 to -4
3N and 3E . . . . .	2	-1.50	-1 to -2
3N and 2E . . . . .	10	-0.40	3 to -3
3N and 2N . . . . .	18	-1.11	1 to -2
3N and 3N . . . . .	10	-1.40	2 to -4

1. In each block of four combinations in Table II, the pattern of frequencies is much the same.
  - a. The frequency for the first combination in each block of four is the least in the block.
  - b. The frequency for the last is next-least in two of the four blocks, tied for next-least in one, and third-least in one.
  - c. That for the second is next-largest in two of the four blocks, tied for next-largest in one, and third-largest in one.
  - d. That for the third is largest in every block.
2. This approximately uniform pattern shows that:
  - a. Unanimity of opinion among judges as to attractive introductions of effective endings is less common than disagreement.
  - b. The quality of writing in introductions was often higher than in endings. The big apples were more likely to be on top in the baskets, as perhaps R. L. Hunt would recommend.
  - c. There was, however, considerable coincidence of quality, or lack of it, in both introductions and endings.

The means of the numerical evaluations of quality of style in the main bodies of the 16 groups of articles listed in Table II suggest that in general the quality in main bodies is inferior to that in either the introductions or endings. The big apples are less likely to be halfway down in the barrels than on top or even on the bottom, as W. T. Foster would probably recommend and as a knowledge of the nature of expository writing might suggest.

But there is somewhat of a tendency for artistic writing in the main bodies of articles to correlate with that in beginnings and endings. The high mean of 1.33 for combination 2A and 3E, the low mean of -1.14 for combination 2N and 3N, the low mean of -1.11 for combination 3N and 2N, and the low mean of -1.40 for combination 3N and 3N all four substantiate such correlation.

That the tendency toward positive correlation, as shown in Table II, is not more pronounced, is mitigated by recalling that N-for either an introduction or an ending may mean either average or poor, whereas quality of writing in main bodies was evaluated in three degrees of excellence, superior, good, and poor.

#### Elements of excellence in writing

There is no satisfactory, suitable, and standardized list of characteristics of good expository writing. One of the objectives of the two surveys reported herewith was to prepare such a list. To this end, the instructions to students in all classes in Research in Education asked the students to rate the style of writing as superior, good, or poor, specified that the evaluating called for analytical thinking, and

that the judges should be clear, be succinct, and be discriminating.

The judges not only evaluated the articles as to quality of style but also enumerated reasons for their decisions.

Twenty characteristics of superior style were coined by the writer, using terms employed by individual judges in their reports on articles regarded by them as being superior. A few actual examples of judges' statements are shown in quotation marks after some of the 20 elements of excellence. The 20 elements of excellence, in descending order of their frequency of mention by the judges in the two surveys, are shown below. (Not more than one tally per judge per element per article was entered in the tabulation, even though an individual judge may have mentioned more characteristics than one which the writer grouped into a single family of similar characteristics.)

#### Master list of elements of excellence in writing

1. Good organization; coherent, orderly, unified, logical, concise, clear-cut, straight-forward, to the point; good continuity and smooth transition from thought to thought; proper length for substance contained. "The ideas tinkle along like a melody."  
"Every sentence gets on with the story."
2. Clearness; simple, definite, direct, precise, comprehensible, easy to understand.
3. Appropriate, simple vocabulary; facile use of well-chosen, euphonious, or poetic words; exact terminology giving tone, imagery, and refinement; alliteration. "Use of slogans and phrases of the day."
4. Apt, adequate, humorous, or effective analogies, allegories, anecdotes, allusions, inferences, examples, illustrations, and figures of speech. "Injection of narrative style into expository material."
5. Pithy, scintillating, well-constructed sentences with variety and change of pace. "Artistically balanced sentences."
6. Forceful, arresting, brilliant, dramatic, brisk, inspiring, challenging, stimulating, enthusiastic, emphatic, original, exciting, invigorating, thought-provoking writing which has zip.
7. Effective use of quotations and documentation; adequate but not too numerous; pertinent; well blended into discourse.
8. Frank, sincere, human, tolerant attitude; use of first person, which makes the reader feel as if the article were written expressly for him. "Gives one the feeling of being an actual participant in the event."
9. Humorous, facetious, subtle, trenchant, incisive, satirical, or sarcastic touches. "The author has humor and trope, not to adorn, but to elucidate."

10. Effective, colorful, vivid, lucid descriptions; pithy definitions; good characterizations. "A revelry in descriptive words and phrases."
11. Forthright, prompt, direct, catchy, fascinating, lively, or impressive introduction or beginning.
12. Well-constructed paragraphs with topic sentences and summarizing sentences.
13. Pointed, convincing, staccato conclusion or summary or ending, with suitable implications.
14. Use of comparisons and sharp contrasts.
15. Use of questions which stimulate interest.
16. Light and carefree style maybe suitable for oral discourse.
17. Appropriate or effective or provocative title.
18. Format; use of half-tones; style of type.
19. Use of italics to emphasize important words or thoughts.
20. Evidence of keen insight into human nature.

Table III shows the frequency with which each of the 20 elements of excellence was listed by judges for articles they considered superior among 15 in the AAUP Bulletin and among 194 other expositions. Not tabulated were such general statements as "A very captivating and interesting style of writing"; "The article grips you and makes you want to finish it"; "The author seems to be a master of a distinctive prose style." Also not tabulated were: (1) "Good English" (Such was taken for granted, although its absence in poor writing was tabulated as a contributing factor to weakness); (2) elements of excellence noted in articles classified as good or poor; (3) comments about content instead of style of writing.

#### Elements of lack of excellence in writing

The procedure for identifying elements of lack of excellence in writing resembled that for identifying elements of excellence. The 50 readers of the 15 articles in the AAUP Bulletin gave reasons why each of the 12 was not among the best written; the teams of judges of the 193 other articles classified some as poorly written and stated why. The elements of lack of excellence were abstracted from the various judges' criticisms and expressed in words found in the same. After some of the elements in the following master list there is an actual quotation, selected because of its expressiveness. If a critic pointed out elements of inexcellence in a good or superior article, it was not counted. Also not tabulated were general statements such as, "It lacked finesse in style."

TABLE III

FREQUENCIES OF MENTION OF ELEMENTS OF EXCELLENCE IN  
SUPERIOR WRITING

Elements of Excellence	Frequencies of Mention		
	In AAUP Bulletin	In Other Articles	Totals
Good organization; smooth transition . . .	91	51	142
Clearness; simplicity . . . . .	62	78	140
Appropriate vocabulary; facile use of words	59	78	137
Apt analogies, examples, figures of speech .	50	54	104
Artistic sentence construction . . . . .	35	58	93
Forceful, dramatic, stimulating style . . . .	21	28	49
Quotations and documentation . . . . .	9	36	45
Frankness; sincerity; humanness . . . . .	24	14	38
Humor; satire; sarcasm; subtlety . . . . .	23	15	38
Effective description and characterization .	--	32	32
Forthright and attractive introductions*	29	--	29
Paragraph construction . . . . .	5	17	22
Pointed, convincing, staccato ending*	11	--	11
Use of comparisons and sharp contrasts . . .	--	8	8
Use of questions effectively . . . . .	--	8	8
Light and carefree style . . . . .	5	1	6
Appropriate or provocative title . . . . .	5	--	5
Format; half-tones; style of type . . . . .	--	2	2
Use of italics for emphasis . . . . .	--	2	2
Keen insight into human nature . . . . .	--	1	1

\*Introductions and endings could not enter in the analyses of the 193 articles covered in the earlier surveys. The judges were instructed to start with the second or third paragraph and to stop one or two paragraphs before the end. Thus, they could not be influenced by the style of writing in the introduction or ending.

Master list of elements of inexcellence in writing

1. Dull, dry, heavy, forced, flat, cold, academic, unimaginative, stilted, tiring, boring, pedestrian, uninspiring, pedantic, strained, unnatural, tedious, formal, mechanical, prosaic, lifeless, plodding, irritating, unstimulating, matter-of-fact style; lacking force, color, punch, humor, vitality, spirit, verve or zest. "I have the feeling the author was as glad to finish writing it as I was to finish reading it." "The tone is made up entirely of flats and is a dull piece of music."
2. Unclear, muddled, obscure, hazy, vague, abstract, indefinite, inexact, not concise, evasive, noncommittal discourse; presumes too much previous knowledge on part of readers; too many modifying elements; insufficient definition of terms. "This article is mumbo-jumbo to me."
3. Long, involved, complicated, or incomplete sentences; too many passive verbs; monotony of sentence structure.
4. Incoherent, choppy, jerky, rambling organization; weak continuity; awkward transitions; loosely knit; depressing development; a hodgepodge of construction; not smooth-flowing; wanders off subject. "Lacks moderation and harmony of parts." "The paragraphs did not reflect orderly thought, but rather seemed to represent a listing of items in the manner of a mail-order-house catalog."
5. Cumbersome, outlandish, trite, stereotyped, monotonous, ostentatious, far-fetched, inadequate, or technical vocabulary; inept or limited word choice; use of slang, profanity, colloquialisms, or irrelevant expressions; wobbly semantics; injection of foreign phrases when English is adequate; flowery phrasing. "Oversized and overstuffed vocabulary." "Clumsy use of fancy adjectives."
6. Wordy, verbose, repetitious, garrulous, redundant writing; too long for substance contained. "Bogged down in verbosity."
7. Inappropriate or unsuccessful emotional appeal; tried to be funny, clever, sarcastic, or satirical but failed; overcrowded with subtleties; weak puns; cynical; strained humor; egotistical; dogmatic, opinionated; overly dramatic; overly sensational; too much like a popular magazine or short story; better suited for oral than for written discourse; tactless; preachy. "The attempts to motivate, to apply 'punch' and embellishments, are crude and ineffective."
8. Too few or too many illustrations, allusions, examples, or references; inept ones; esoteric ones; trite analogies or figures of speech. "Too few illustrations or personalized episodes to enable one to relive the experience."
9. Poorly constructed paragraphs or improper division into paragraphs; monotony of types of paragraphs. "The author divided his article into paragraphs as though he were slicing bread."

10. Inadequate or superficial treatment; covered too much territory, subject too big for the treatment it got.
11. Errors in grammar or punctuation.
12. Unappealing introduction or opening paragraph; slow or confused in getting started.
13. Too many or too few or too weak quotations or dependence on authority or use of documentation, or use of quotation marks around words which are not quoted (usually too many.)
14. Abrupt, weak, disappointing, unimpressive, or inconclusive endings or closing paragraphs; no conclusion.
15. Misleading or inappropriate title.
16. Use of first or second person too prevalent, unappealing, or annoying. "Two pages contained nineteen I's".
17. Insincerity; tries to appear scholarly; pedantic.
18. Too frequent use of ineffective or inappropriate use of questions.
19. Format or typography.
20. Use of outline form.
21. Comparisons which were inappropriate or out of place.
22. Use of italics for emphasis.

Table IV shows the frequency of mention of the 22 elements of lack of excellence by the several judges in their criticisms of the two groups of expositions.

#### Comparison of elements of excellence and of inexcellence

Obviously, and probably as should be expected, the elements of in-excellence in Table IV, and even their rank order in the table, are much the same as the opposites of the elements of excellence in Table III. The highest-ranking exception to this is wordiness or verbosity, which has no exact opposite.

The number of elements of lack of excellence, the number of ideas annexed to most of them, and the average number of frequencies for each, were greater than for elements of excellence. These differences are accounted for in part by there being more articles in each set judged poor than judged superior. But a more potent factor probably is the so-called halo effect. Several surveys by the writer, and other earlier researches, have shown that elements of weakness are more identifiable than elements of strength.

If something is good, both casual observers and expert analysts are less able to show diagnostically wherein it is good than they are to show wherein something bad is bad. The presence of errors in grammar and punctuation among elements of lack of excellence is an example in point. Good English in superior writing is taken for granted.

TABLE IV

## FREQUENCIES OF MENTION OF INEXCELLENCE IN POOR WRITING

Elements of Inexcellence	Frequencies of Mention		
	In AAUP Bulletin	In Other Articles	Totals
Dull, dry style; lack of verve . . . . .	137	101	238
Unclear, muddled, evasive writing . . . . .	135	83	218
Weak sentence structure . . . . .	80	120	200
Incoherent, rambling organization . . . . .	112	49	161
Cumbersome, trite, inept vocabulary . . . . .	93	66	159
Wordy, verbose, redundant writing . . . . .	50	30	80
Inappropriate emotional appeal . . . . .	42	17	59
Inept analogies and figures of speech . . . . .	36	11	47
Poor paragraph construction . . . . .	10	35	45
Inadequate or superficial treatment . . . . .	15	21	36
Errors in grammar and punctuation . . . . .	7	29	36
Unappealing introductions* . . . . .	32	--	32
Use of quotations and documentation . . . . .	17	11	28
Abrupt, unimpressive ending* . . . . .	24	--	24
Misleading or inappropriate title . . . . .	10	2	12
Use of first or second person . . . . .	5	6	11
Insincerity; pedantry . . . . .	--	6	6
Ineffective use of questions . . . . .	3	2	5
Format or typography . . . . .	--	3	3
Use of outline form . . . . .	--	3	3
Inappropriate comparisons . . . . .	--	1	1
Use of italics for emphasis . . . . .	--	1	1

\*Introductions and endings could not here, just as in Table III.

### Education periodicals versus other periodicals

Except for a single sample from the NEA Journal, the 193 articles selected from those covered in the two earlier surveys in this series represent five periodicals: American Historical Review, 55; Journal of the American Association for Health, Physical Education, and Recreation, 50; American School Board Journal, 36; Journal of Educational Research, 29; and House Beautiful, 22.

Counting a rating of superior as +1, a rating of good as 0, and a rating of poor as -1, the numerical evaluation of the merit of writing in the main body of an article, as judged by the five separate graduate students, could run as high as 5 or as low as -5. The mean scores for the articles in the five periodicals listed above are 0.00, -0.46, -0.61, and -1.41, and 0.14, respectively. The mean score for American School Board Journal and Journal of Educational Research together is 0.97, while the mean for the noneducation periodicals is -0.16. This does not speak well for education.

### Summary and conclusions

Earlier articles in this series on art in writing for educational periodicals dealt with the introduction and the ending, while this concluding report concentrates on the longer middle portion, or main body.

Evidence indicates that the judges' evaluation of merit were fairly reliable.

The quality of writing in the main bodies of magazine articles is not so meritorious as in the introductions or even in the endings. There is some degree of correlation between quality of style in the main bodies, however, and quality of style at beginnings and ends.

Elements of inexcitement in writing are more identifiable than elements of excellence. A score or more of elements of each type were abstracted from the judges' criticisms. On the whole, the elements of inexcitement are opposites of elements of excellence. The top five elements of excellence are: coherent organization, clearness, well-chosen vocabulary, apt analogies and figures of speech, and artistic sentence construction. The top five elements of lack of excellence are: dullness, unclarity, weak or monotonous sentence structure, incoherent organization, and cumbersome vocabulary.

It may be that the art of writing can be learned but not taught. In any case, whatever learning or teaching occurs must be by elements of the art of writing and not by the art as a whole. It was for this reason that this series of surveys was made and published.

Writers for educational periodicals are particularly in need of reform. The earlier surveys in this series found contributors to education magazines less apt at the art of writing than contributors to other

periodicals. The present report repeats the same sad story.

The evidence against educationists is too consistent and too convincing to be shrugged off as inconsequential. Educationists are guilty! Guilty of what? Either of incompetence or of indifference. If they are not too amental to learn, they dare not be too apathetical. Just as none is so blind as he who will not see, so is none so ignorant as he who will not learn. How much longer will writers for educational periodicals remain willing to see themselves outclassed?

# IS GENERAL MATHEMATICS OR ALGEBRA PROVIDING GREATER OPPORTUNITY TO ATTAIN THE RECOMMENDED MATHEMATICAL COMPETENCIES?\*

MILTON W. BECKMANN  
University of Nebraska  
Lincoln, Nebraska

IN 1944, THE board of directors of the National Council of Teachers of Mathematics created a Commission on Post-War Plans. Its assignment was to decide the basic mathematics needed for personal use by the post-war citizen... not just for those who planned to go to college or for those engaged in technical professions and trades, but for everyone. On the basis of their deliberations, they were to prescribe realistic goals in the post-war teaching of mathematics.

Thesis Number 1, headlining the Commission's various reports, was mathematical literacy for all who can possibly achieve it. Every teacher of mathematics from grade one through junior college is vitally concerned with this thesis. Other recommendations followed: the large high school should provide a double track in mathematics; the teaching of arithmetic can be and should be improved, we should differentiate on the basis of needs; the sequential courses should be greatly improved; and so forth. But this primary aim —mathematical literacy for all — was regarded throughout as the most important job facing the Commission.

## Mathematical Literacy Defined

How much mathematics is needed by every citizen? How much mathematics does it take to be "mathematically literate"... to fit Mr. Average Citizen culturally and vocationally to his modern, scientifically charged environment? The essentials for functional competence in mathematics were placed as questions in a Check List by the Post-War Commission. If the 29 questions in the Check List<sup>1</sup> can be answered with a "yes" the student can feel reasonably competent when it comes to dealing with the problems of everyday affairs.

\* The Level of Mathematical Competencies And Relative Gains In Competency Of Pupils Enrolled In Algebra And General Mathematics. Unpublished Ph.D. dissertation, University of Nebraska, 1951.

1. Raleigh Schooling (Chairman), "The Second Report of the Commission on Post-War Plans," The Mathematics Teacher, 38:197-198, (May, 1945)

Competence in handling the kit of tools given above is very useful in solving the great majority of the average man's problems. It is the basic mathematics needed in a great many vocations. The Check List of 29 competencies suggests that competence in mathematics is almost as crucial as literacy in communication. Therefore, the Commission held that the school must guarantee competence in these specific matters, to all students who can possibly achieve it.

### Providing Mathematical Literacy

What mathematics are we going to require so that all pupils who graduate from our high schools will become "mathematically literate"? In some schools a few of the exceptionally capable students will have achieved competence in these essentials by the time they have completed the eighth grade. Possibly some of the items could be learned in a course in business arithmetic, however, pupils could learn many of the same things in a one year course in general mathematics. They would achieve still more of them by taking two years of general mathematics. Many schools are now teaching a course called consumer mathematics. Again many of these competencies will be taught in this course.

Those students who take four years of the usual sequential courses in mathematics will obtain most of the competencies. However, there would be, no doubt, many gaps in the traditional courses as now taught. If one were planning to become competent mathematically in only this Check List, the old traditional road might prove quite expensive as far as effort and time are concerned. It might be better to take some of the general courses described.

The writer has felt for a long time that there was a need for research attacking the problem. Many articles have been written in an attempt to make the general mathematics courses respectable and desirable. Much had been written concerning the revision of the traditional mathematic courses. Several excellent mathematics bulletins have been developed by statewide committees. One of these bulletins issued by the State Superintendent of Wisconsin, serves as a guide for a sequence in general mathematics. These articles, reports, and bulletins have been stimulating. However, very little research has attempted to discover what courses contribute most towards mathematical literacy.

### Statement of the Problem

Evidence now discloses the rapid elimination of algebra and the substitution of applied or general mathematics in Nebraska High Schools. This study presents factual data to reveal whether students enrolled

in general mathematics or first year algebra are provided the greater opportunity to attain the recommended mathematical competencies for general education.

### Organization and Plan of the Study

To do this, it was necessary to find schools that required all ninth grade pupils to take algebra, and schools that required all ninth grade pupils to take general mathematics. This information was found in the office of the State Department of Public Instruction. Their records showed what courses schools had offered for the school year 1949-1950. Approximately one-hundred schools met this criterion. It must be remembered that a number of high schools in Nebraska offer both algebra and general mathematics in the ninth grade. These schools could not be included since the assumption of homogeneity to the two types of classes would have been unwarranted. Out of the one hundred schools listed, fifty-one were selected for inclusion in this study. In choosing the fifty-one schools, consideration was given to geographical location so that all areas of Nebraska are represented.

Early in September, 1950, letters were written to the superintendents of the 51 schools requesting their participation in the study. Only one declined to participate in the study. Three schools could not be included because they had changed their course offering the preceding year. Thus 47 schools were included originally in the study. Not all of the schools which agreed to participate were able to complete the study. Forty-two Nebraska High Schools, with twelve hundred and ninety-six students, were selected for final inclusion in this study. Twenty-five of these schools, with 734 pupils participating, required all ninth graders to take general mathematics. Seventeen schools, with 562 pupils participating required all ninth graders to take algebra.

### Evaluative Instruments

In planning the study, it was decided to use a mathematics test which would measure functional competence in mathematics in terms of the 29 competencies found in the Check List. Since no satisfactory measuring instrument could be found, it was decided that one should be constructed. A detailed explanation of the construction of the Mathematical Literacy Test is given in Chapter V of a study conducted by the writer.<sup>2</sup> The steps taken in this study follow rather closely the usual procedure in test construction.

2. Milton W. Beckmann. op. cit.

It was also necessary to select a test to measure intelligence. For this purpose, the test selected was the New California Short-Form Test of Mental Maturity.

### Administration of Tests

The superintendents were told that participation in the study would involve the following commitments on their part:

1. To give a mathematical literacy test to all of their ninth grade pupils. These tests were to be given early in the school year. The time needed would be approximately two forty-minute periods.
2. To give intelligence tests to all of their ninth grade pupils sometime in February or whenever it would be most convenient.
3. To give a second mathematics test to all ninth grade pupils toward the end of the school year. The time needed would be approximately two forty-minute periods.

They were to continue their work in mathematics as they had planned, making no changes in content or methods because of participation in the study.

The schools selected had ninth grade classes with enrollments ranging from eight to 147.

### Analysis of Covariance of Test Scores

The Mathematical Literacy Test was made up of 110 items. One was discarded because of a typographical error. The highest possible score that could be attained was 109.

The initial mean score on the Mathematical Literacy Test for 562 algebra students is 46.54, and the final mean score is 54.44. This was a gain of 7.90 in the means of the algebra group. Similarly the means of the general mathematics group were 44.89 and 54.04 respectively. This was an increase of 9.15. One of the important findings of this study was this astonishingly low gain in understanding and mastery of the 29 competencies, both in algebra and general mathematics after a year of study.

The difference between the obtained means of the algebra group and the general mathematics group is 1.25 in favor of the latter. Is this difference significant? That is to say, would further testing of similar samples of general mathematics and algebra students give virtually the same result; or is it probable that the mean difference would be reduced to zero, or even reversed in favor of the algebra students? The

method<sup>3</sup> of analysis of covariance will enable us to answer this question.

The actual difference between the gain in means of the algebra and the general mathematics groups from the initial to the final test is 1.25. Let us assume that the difference between the population means of the two groups is zero, and that except for accidental errors mean differences from sample to sample would all be zero. In making this assumption we are setting up the null hypothesis.

The null hypothesis. A better rule-of-thumb procedure for testing the statistical significance of a difference is to set up the null hypothesis that there is no difference between the two population or universe measures (parameters).<sup>4</sup>

As a criterion the gains made by the students on an initial and final testing of a mathematical literacy test were used. Since scholastic aptitude could conceivably influence each student's response to the criterion, these individual differences were controlled by obtaining intelligence quotients as a measure of scholastic aptitude. By using these intelligence quotients on the control variable in the analysis of covariance, the possible bias introduced by individual differences will be removed in so far as these factors adequately represent the differences in question. The sums and means of the criterion and the control variable are shown in Table I.

A test of significance can now be made of the null hypothesis that students who were required to take algebra in the ninth grade do not differ in gain in achievement on the competencies as proposed by the Post-War Commission from those ninth grade students who were required to take general mathematics with their intelligence quotients controlled. The analysis of covariance is shown on Table II.

The F-value of 19.2490 with 1 and 390 degrees of freedom is significant beyond the 1% level of confidence. Therefore, when the criterion means of the two subgroups are adjusted for individual differences in scholastic aptitude, the difference between the mean difference on achievement of students on the competencies of those who were required to take general mathematics, is so large that it undoubtedly was not caused by a sampling accident. Presumably the difference can be attributed to the fact that the competencies are better taught in the general mathematics classes.

There is a possibility that other essential concepts, principles, and processes are being taught in classes of general mathematics and algebra which are not included in the Commission's Check List, but this

3. J.S. Ahman and others. "Analysis of Covariance," (mimeographed).

4. Quinn McNemar, Psychological Statistics, (New York: John Wiley & Sons, Inc., 1949), p. 65.

TABLE I

SUMS AND MEANS OF THE CRITERION AND THE CONTROL VARIABLE  
FOR NINTH GRADE MATHEMATICS STUDENTS

Group	Number	Growth on Mathematical Literacy Test		Intelligence Quotient	
		$\Sigma Y$	$\bar{Y}$	$\Sigma X^*$	$\bar{X}$
Gen. Math.	734	6713	9.15	4883	6.65
Algebra	562	4439	7.90	5041	8.96
Total	1296	11152	8.60	9924	7.13

\*Scores were "reduced" by the subtraction of 100.

TABLE II

TEST OF SIGNIFICANCE OF INFLUENCE OF ALGEBRA AND GENERAL  
MATHEMATICS ON GAINS IN ACHIEVEMENT OF THE COMPETENCIES  
IN NINTH GRADE MATHEMATICS

Source of Variation	Degrees of Freedom	Residuals	
		Sum of Squares	Mean Square
Total	1294	99,605.3644	
Within Subgroups	1293	98,144.2811	75.9034
Difference	1	1,461.0833	1,461.0833
$F, 1293 - \frac{1461.0833}{75.9034} = 19.2490$		$t = \sqrt{F} = \sqrt{19.2490} = 4.38$	

does not detract from the fact that the relative gains in understanding and mastery of the 29 competencies, both in algebra and general mathematics, are low, and that the competencies are better taught in the general mathematics classes.

### Implications

The Report of the Commission on Post-War Plans has been widely approved and supported by authorities in the field of mathematics teaching and by educators generally. If we accept the competencies as valid objectives for mathematics in the schools, then the low level of mathematical competency and the relatively small gains made by ninth grade students enrolled in algebra and general mathematics in Nebraska, indicated by scores on this test of mathematical literacy, points out that one or more of the following conditions apply:

1. That our present curriculum in mathematics is poorly planned and developed.
2. That the teaching of mathematics is poorly done in our schools, or
3. That the competencies as proposed by the Post-War Commission have not been accepted a valid objectives by those planning the curriculum of the schools.

The writer believes that our present curriculum is poorly planned and developed and responsible for a great deal of our mathematical illiteracy.

The level of mathematical attainment is astonishingly low after students have spent a year in the study of algebra or general mathematics. One cannot help but speculate what the situation must be like in schools where no mathematics is required. Students entering and leaving the ninth grade should be required to take more mathematics, the kind which will assure achievement of the essentials for functional competence in mathematics. After the student has acquired the minimum essentials, he then should be offered an opportunity for further mathematical training.

# SOCIAL ACCEPTABILITY AND SOCIAL REJECTION OF THE UNDERAGE, AT-AGE AND OVERAGE PUPILS IN THE SIXTH GRADE

VAGHARSH H. BEDOIAN\*  
Sherman Oaks, California

FOR THE purposes of this study social acceptability was determined by the combined weighted scores of a four criteria sociometric test. The test actually had five criteria, but one criterion was used to determine social rejection. The four criteria were as follows: (1) If you were changed from this class to another class which boy or girl from this class would you choose to go with you? (2) Which boy or girl from this class would you choose as a team captain for a game of sports? (3) Which boy or girl from this class would you choose as the president of your class? (4) Which boy or girl from this class would you choose to help you if you needed help to do your school work?

A total of 743 pupils from 22 classes was allowed three choices for each question. The choices were weighted as follows: Five points were given for the first choice. Three points were allowed for a second choice and one point for a third choice. The scores were tabulated and standard scores computed using the dual-deviation technique developed by Whitson<sup>1</sup> for this type of data. It is interesting to note that out of 743 cases studied only 229 children were above the theoretical mean while 514 fell below the theoretical mean. This shows that thirty-one per cent of the cases received most of the first and second choices while sixty-nine per cent of the cases either were not chosen or they received combined weighted scores of less than thirty-six which was the theoretical mean.

The theoretical mean was the individual score that all pupils within a classroom would receive if all of them received the same score on the sociometric questionnaire. The theoretical mean may also be considered the mean of the raw scores assigned on a chance basis. The theoretical mean was computed according to directions in Milo Whitson's study.

\*14746 Greenleaf Street.

1. Milo Whitson, Statistical and Geometric Techniques for Sociometric Data. Unpublished Doctor's Dissertation, University of Southern California, Los Angeles, Calif., June, 1949.

Social Acceptability

Mean social acceptability raw scores of the underage, at-age, and overage pupils were compared. Underageness and overageness were determined by taking those who were nine months below or above the mean chronological age of their half-grade placement. Because the groups were not comparable the raw scores of the entire distribution could not be combined and averaged. However, an examination of the data shows the basic trend. In ten out of twenty-two classes the underage pupils earned the highest mean score over all competitors. In five of the remaining twelve classes the mean of the underage pupils was higher than the overage pupils. The underage pupils were excelled by the at-age pupils in only ten out of twenty-two classes. In five classes the overage pupils received a higher mean score than the underage pupils and in only four classes did the overage pupils receive higher mean scores than the underage and at-age pupils.

The trend then seems clear. The underage pupils have made the best showing as a group. The at-age pupils have held a position in the middle while the overage pupils have made the poorer showing. It is interesting to study how the overage pupils received higher average social acceptability scores in some of the groups. A Central-American child was the only overage pupil in class twelve. His principal said the boy was an excellent athlete and very popular with his classmates. This becomes evident when one finds that he earned the highest sports leadership score. He received most of his points there. This fact stresses the importance of choosing the proper questions for a sociometric questionnaire. If this type of question had been omitted, the boy would have been considered an "isolate" on the basis of the data.

School ten, class seventeen, presented a different reason for the overage pupils receiving higher scores. In this particular school the sixth grade is divided into two separate classes, an accelerated group and a slow group. In the slow group there were nine overage pupils, the highest number of the twenty-two groups studied. The two underage pupils have fallen by the wayside while the at-age pupils assume a position between the underage and overage pupils. In class seventeen the youngest child received the highest rejection score.

School ten, class eighteen, which is the advanced group mentioned above presents a picture almost opposite to its sister class. In class eighteen the underage pupils made the best showing by receiving the highest mean scores while the overage pupils made the poorest showing. These data present some evidence that the overage pupils enjoy greater recognition and status within a class when they are placed in a class whose mean age approximate their own chronological age.

In a further analysis of the problem the means of the social acceptability standard scores were compared. In fifteen out of twenty-two

classes a consistent relationship was shown among the underage, at-age and overage pupils. The underage pupils received the smallest negative standard scores, the at-age pupils the next highest and the overage pupils possessed the highest negative standard scores. A high negative standard score would indicate a low social acceptability raw score. In two of the remaining seven classes the underage pupils earned lower negative standard scores than the overage pupils. In five out of twenty-two classes studied the overage pupils received lower negative standard scores than the underage and at-age pupils. In other words the overage pupils earned the lowest social acceptability raw scores in all but five classes.

The present data indicate that the retarded child lacks social acceptance in his group. He is denied the pleasure of status and recognition within the social framework of the classroom.

The differences, among the three groups are quite significant. The difference between the overage<sup>2</sup> pupils and the underage pupils was -1.04 with a critical ratio of -3.9 which is significant beyond the one per cent level. Mean differences of the overage and at-age pupils was -.71 with a critical ratio of -3.1 which is significant beyond the one per cent level. The difference between the at-age and the underage pupils was -.33 with a critical ratio of -2.0 which is significant at the five per cent level of confidence.

An attempt to help the slow child results in further retardation, and the popularity of the child suffers. The child lacks recognition and status in the group. If the aim of education is the total development of the child, an important aspect of the child's development is certainly missing.

A further analysis of social acceptability was made by studying the place held by the underage, at-age and overage pupils in the extremes of popularity, this is, the "stars of attraction." The "stars of attraction" are those pupils who received the highest scores in social acceptability. For the purposes of this research the four highest score recipients in each class were studied. Twenty-three out of 188 or 12.2 per cent underage pupils were among the four highest "stars of attraction". Sixty-three out of 505 at-age pupils or 12.5 per cent were among the "stars of attraction". Two out of fifty or four per cent of the overage pupils were among the four highest "stars" in the various classes studied. This clearly shows the advantage the underage and at-age pupils have over the overage pupils. The underage pupil is on a par with the at-age pupil, but the overage pupil is almost out of the picture. The difference between the means of the underage and the overage pupils was 8.2 per

2. In presenting the differences of means, the first-mentioned is always the greater of the two. This scheme is followed throughout the research.

cent with a critical ratio of 2.2 which is significant beyond the five per cent level. The mean difference between the at-age and overage pupils was 8.5 per cent with a critical ratio of 2.7, which is significant beyond the one per cent level of confidence. A statistically insignificant mean difference of .30 was found between the at-age and underage pupils.

Because certain underage pupils were rejected by their peers an effort was made to determine the relationship between social acceptability and the degree of underageness. The underage pupils were divided into three groups: (1) those who were underage nine to eleven months, (2) those who were twelve to fourteen months underage, (3) those who were underage fifteen months and more. The nine to twelve months underage pupils had the lowest negative mean, that is, this group of underage pupils was socially more acceptable than the other two groups of underage pupils. The twelve to fourteen months underage pupils received the highest negative mean. The difference between the means of the nine to eleven months underage and the twelve to fourteen months underage pupils was -.83 with a critical ratio of 2.6 which is significant at the one per cent level. Other differences were not significant.

From these data it appears that slightly underage pupils have the advantage over the extremely underage pupils within a given group. Those who are underage from twelve to fourteen months are not socially attractive as the slightly underage pupils. On the other hand, the pupils who are younger by fifteen months show a little gain in social acceptance over the eleven to fourteen months underage pupils. Although the extremely underage pupils have less social acceptability than the slightly underage children, they still have the advantage and excel in popularity over the at-age and overage pupils by a considerable margin.

### Social Rejection

Along with the sociometric test an additional question was asked to determine social rejection: If you were changed from this class to another class which boy or girl would you not choose? Three choices were allowed and the choices were weighted in the same manner as the rest of the sociometric questionnaire. In sociometric literature the word "isolate" is used to designate the person who is not chosen or is chosen very seldom by his peers. The word "rejectee", however, is used to designate one who is actively disliked by his peers. Comparisons were made of the positions underage, at-age and overage pupils held among the four highest "rejectees", that is, the pupils receiving the highest score for the above question in their class. In fourteen classes the largest percentage of placement among the four highest "rejectees" came from the overage pupils. Of the remaining eight classes, two classes did not have any overage pupils. The underage

pupils comprised the largest percentage among the four highest "rejectees" in only three classes. The at-age pupils composed the largest percentage in five of the classes.

Seventeen out of 188 underage pupils or nine per cent were among the four highest "rejectees". Fifty-three out of 505 at-age pupils or 10.5 per cent were among the four highest "rejectees". Eighteen out of fifty overage pupils or thirty-six per cent were among the four highest "rejectees". A mean difference of twenty-seven per cent was found between the overage and underage pupils with a critical ratio of 3.8 which is significant beyond the one per cent level. The difference between the means of the overage and at-age pupils was 25.5 per cent with a critical ratio of 3.7 which is significant beyond the one per cent level of confidence. The difference between the means of the at-age and underage was not significant.

In the above analysis the overage pupils received the highest scores, but unfortunately the highest scores represent the least desirable pupil reactions. The overage pupils appear to be the least popular within the class, the at-age pupils are next while the underage pupils seem to be the most popular pupils within the group. Class seventeen, which is a "slow" group, had the highest number of overage pupils in any of the classes studied, yet none of the overage pupils were among the four highest "rejectees". Fifty per cent of the underage pupils in class seventeen, however, were among the four highest "rejectees". On the other hand, in class eighteen the "advanced" group in school ten, fifty per cent of the overage pupils were among the four highest "rejectees", but only ten per cent of the underage pupils were among the four highest "rejectees".

Social rejection raw score averages for the underage, at-age and overage pupils were studied. Because these scores had not been converted into standard scores the entire distribution could not be combined so as to obtain a mean. The trend, however, is apparent. In sixteen out of twenty-two classes studied the overage pupils received the highest rejection scores. Two classes had no overage pupils. In three classes the at-age pupils possessed the highest average rejection scores. The underage pupils had the highest average rejection score in but one class and that was class seventeen, which had the greatest number of overage pupils of any class. In fact, a girl, the youngest in class seventeen, possessed the highest rejection score and no acceptance choices. This surely is in contrast to a boy in class one, the youngest pupil in his class, who received the highest acceptance score in class one. In class eighteen the situation was reversed since there were only two overage pupils. In class eighteen the overage pupils received a high average rejection score whereas the underage pupils received a low average rejection score. It was pointed out that the mean age of class seventeen, the "slow" group in school ten, was twelve years and

four months and the mean age of class eighteen, the "advanced" group in school ten, was eleven years and two months. When the overage pupils are in a group of fellow overage pupils they in turn have social status and recognition. When the overage pupils constitute a minority group in the class they become an object of scorn unless they possess some special asset such as outstanding sports ability.

### Summary

In the present research social acceptance and social rejection data were presented. Comparisons were made and significant differences noted among the underage, at-age and overage pupils. Special study was also made of the effect of the degree of underageness on social acceptability.

1. The underage pupils received significantly higher social acceptance scores while the overage pupils' scores were significantly lower than the underage and at-age pupils. Pupils only slightly underage appear to enjoy higher status than pupils who are underage to a greater degree. It is important to note that in spite of the loss of acceptability by the extremely underage pupils they still excel the at-age and overage children. It is evident from the present data that age determines social acceptance to a considerable degree. The data from class seventeen clearly show that if the overage pupils are placed together they enjoy a higher degree of status than the overage pupils who do not make up a large share of the class.

2. The overage pupils appear to possess significantly higher rejection scores than the at-age or underage pupils. The data indicate that not only are the overage ignored in classroom activities, but that they are actively disliked by their classmates. The individual is rare who is not affected when he is ignored and relegated to a life of isolation, thus suffering a loss of prestige within his group.

3. A significantly larger proportion of underage pupils were among the "stars of attraction" — those who received the highest acceptance scores. On the other hand, the overage pupils constituted a significantly larger proportion of the four highest "rejectees" than the underage pupils.

## BIBLIOGRAPHY

- Bedoian, Vagharsh H., "Sociometric Status and Mental Health of Over-age and Underage Pupils in the Sixth Grade." Unpublished Doctor's Dissertation, University of Southern California, Los Angeles, 1940. 169 pp.
- Bonney, Merl E., Popular and Unpopular Children, A Sociometric Study. Sociometry Monographs, No. 9. New York Beacon House, 1947. 81 pp.
- \_\_\_\_\_, "A Study of the Sociometric Process Among Sixth Grade Children," Journal of Educational Psychology, 37:359-372, 1946.
- Farley, Eugene S., "The Influence of Grading and Promotional Policies Upon Pupil Development," Elementary School Principal, 16:268-274, July, 1937.
- \_\_\_\_\_, "Regarding Repeaters; Sad Effects of Failure Upon The Child," The Nation's Schools, 18:37-39, October, 1936.
- Heintz, Emil, "Adjustment Problems of Class Status," Phi Delta Kappan, 30:290-293, April, 1949.
- Jennings, Helen, Leadership and Isolation. New York: Longmans, Green and Co., 1934, 240 pp.
- \_\_\_\_\_, Sociometry in Group Relations. Washington, D. C., American Council on Education, 1948. 85 pp.
- Kuhlen, Raymond G. and Howard S. Bretschi, "Sociometric Status and Personal Problems of Adolescents," Sociometry, 10:122-132, May, 1947.
- Moreno, Jacob L., Who Shall Survive? Washington, D. C.: Nervous and Mental Disease Publishing Co., 1934. 450 pp.
- Moreno, Jacob L., Helen Jennings and Richard Stockton, "Sociometry in the Classroom," Sociometry, 6:425-428, November, 1943.
- Northway, Marl L., "Outsiders: A Study of the Personality Factors of Children Least Acceptable to Their Age-Mates," Sociometry, 7:10-25, February, 1944.
- \_\_\_\_\_, "Social Acceptability Test," Sociometry, 5:180-184, May, 1942.

Sandin, Adolph, Social and Emotional Adjustments of Regularly Promoted and Non-Promoted Pupils, Child Development Monographs, No. 32. New York: Bureau of Publications, Teachers' College, Columbia University, 1944, 142 pp.

Whitson, Milo E., Statistical and Geometric Techniques for Sociometric Data. Unpublished Doctor's Dissertation, University of Southern California, Los Angeles, June, 1944. 233 pp.

# RESEARCH IN NEGRO COLLEGES

J. IRVING E. SCOTT\*  
State A. and M. College  
Orangeburg, South Carolina

TODAY, MORE than ever, there is a growing consciousness that one of the functions of every teacher is research. Clinics, work-shops, conferences and institutes are conducted frequently to encourage and assist teachers to improve their efficiency and to avail themselves of community resources through research. This is particularly true in institutions of higher learning. The worth of research to the overall program of these institutions is clearly brought out by Buckingham:

In colleges and universities, faculty members are appointed, promoted, and paid not only because they can teach but also because they can conduct original investigations. 1

In speaking of the importance of research, Buckingham says, "The research activities of faculty members give vitality to their teaching." 2 In the light of the foregoing, it is important to know how much research is going on in Negro colleges and whether or not these institutions are carrying on an organized program of research.

In an effort to secure this information there were certain definite things that the author thought might lead to an appraisal of the program. The basic questions selected to obtain this information were:

1. Has the college a research committee?
2. Has the college a director of research?
3. What are the functions of the research committee?
4. What are the functions of the director of research?
5. Is there a research bulletin published by the college?
6. Is there a section of the school bulletin or catalog which is given to faculty members for research?
7. Is there an independent research bulletin published at the college?
8. How many faculty members have written books within the last five years?
9. What financial help is given faculty members for research?
10. Is there any reduction in the teaching load of people who do research?

\*Editor-in-Chief of Negro Educational Review.

1. Burdette Ross Buckingham, Research For Teachers. (New York: Silver, Burdett and Company, 1926), p. 379.

2. Ibid., p. 380

It would appear that a questionnaire for the information sought would take into consideration the many faculty members who have written articles for educational journals. This question was intentionally left out because of the large number of articles written that by no stretch of imagination could be considered as the result of research.

Questionnaires were sent to 103 institutions of higher learning for Negroes. There were responses from 69 of these institutions of varying sizes. These 69 institutions have a combined teaching staff of 3401 teachers. Of the 69 institutions studied 20 were public supported and 49 were privately supported. Included in the study are 11 of the 17 land-grant colleges and 13 of the 18 Negro institutions that offer graduate work.

From Table I it will be noted that 26 of the 69 institutions responding have research committees while 43 have no such committees. A breakdown according to the size of the school indicates that of the 14 schools with 20 and less teachers only three or 27% have research committees. In the schools with 20 to 50 teachers 10 or 33% have research committees whereas in the schools with 51 to 100 and over 100 teachers, about 50% have research committees. This clearly indicates that the larger the school the better the chances for having a research committee.

Of much interest, however, is the nature of the functions of these research committees. The following are the functions listed by the respondents:

1. To stimulate and develop publications of research.
2. To develop curriculum divisions and foster revision of curriculum in light of student success.
3. To coordinate all research activities.
4. To coordinate the research studies of the graduate division.
5. To deal with master's theses and conduct institutional research.
6. To approve projects and requests for financial assistance.
7. To evaluate all research work and special studies.
8. To encourage research among the members of the staff, to serve as advisors and to recommend financial, secretarial and academic assistance.
9. To study local problems.

#### Director of Research

Of the 69 institutions responding to the questionnaire only 4 have a Director of Research. These institutions are Langston University, Hampton Institute, Prairie View College and Wilberforce University. The functions of these directors are:

TABLE I

NEGRO INSTITUTIONS OF HIGHER LEARNING HAVING RESEARCH  
COMMITTEES AND RESEARCH DIRECTORS

Size of School	Number of Institutions Responding	Research Committees		Research Directors	
		Yes	No	Yes	No
20 and less teachers	14	3	11	-	13
20-50 teachers	30	10	20	2	26
51-100 teachers	17	9	8	2	12
Over 100 teachers	8	4	4	-	8
and Total	69	26	43	4	59

1. To coordinate the program of research in the schools.
2. To stimulate research.
3. To supervise and coordinate activities of research.

As will be observed the functions of the director of research are very similar to those of the research committee.

#### Research Bulletins

The emphasis that is placed on research in Negro colleges can be seen from the number of colleges having committees on research and directors of research. Further information is had from the research bulletins that are published by Negro colleges. Of the 69 institutions responding to the questionnaire only two institutions, Florida A. and M. College and Langston University, publish research bulletins. Ten of these institutions however, give some pages of the school bulletin or catalog for the publication of research material of faculty members. The Quarterly Review of Higher Education published at Johnson C. Smith University, The Negro Educational Review published at South Carolina State College, and the Journal of Negro Education published at Howard University are all independent magazines.

#### Faculty Members Writing Books

The question on how many faculty members have written books since 1945 was much misunderstood. For this the author takes the blame since a more definite wording would have brought out less confusing answers. The list in Table II gives the status as of this writing.

#### Aids Given to Research Workers

It must be admitted that despite one's willingness to carry on research, there are certain factors that will help or hinder one's effectiveness. A number of known research men were questioned as to what they consider the greatest handicap in carrying on research in Negro colleges and unanimously they agreed:

1. Lack of finance.
2. Lack of secretarial help.
3. Lack of time and facilities.

Keeping these things in mind it was important to find out what is being done in Negro colleges to relieve the situation given above. Table III gives some of this information.

TABLE II

## INSTITUTIONS WHOSE FACULTY MEMBERS HAVE WRITTEN BOOKS SINCE 1945

Institutions	No. of Books Since 1945	No. Faculty Members Written Bks.	Names of Books Written
South Carolina State College	4	3	1. Guide to Thesis Writing 2. Inside Germany 3. Negro Students and Their Colleges 4. Finding My Way
Stowe Teacher's College	3	2	1. American Literature by Negro Authors 2. The Negro Newspaper 3. The Negro's Adventure in General Business
Southern University	3	3	1. Experimental Food Biochemistry for Students of Home Economics and Agriculture 2. Plant Biology Manual 3. Simplified Tailoring Manual
Tuskegee Institute	5	4	1. Mathematics for Students of General Education 2. Negro Year Book 3. General Education in the Negro College 4. Africa Advancing 5. The School Comes to the Farm
Hampton Institute	2	2	1. Negro Voices in American Fiction 2. They Came in Chains 3. Strangers and Alone
Philander Smith	1	1	1. Say Amen Brother
Arkansas State College	1	1	1. The Changed Political Thought of the Negro, 1915-1940.
Wilberforce University	3	3	1. A Short History of Philosophy from the Renaissance to Hegel 2. The Young Scholar 3. Po Polska Bron (To Fight On)
Lincoln University	1	1	1. Caste, Class, and Race
Atlanta University	3	3	1. Practical Farming for the South 2. Planning 3. Negro Business and Business Education.
Morehouse College	3	2	1. Seeking To Be Christian in Race Relations 2. A Gospel for The Social Awakening 3. The Brotherhood of Sleeping Car Porters

TABLE III  
AIDS GRANTED RESEARCH WORKERS IN NEGRO COLLEGES

Types of Aids	Number of Institutions
1. The Carnegie Grants	6
2. G. E. B. Grants	1
3. Some travel funds	2
4. Reasonable time off from teaching	1
5. One quarter off with full pay	1
6. Small temporary grants	3
7. Grants from experimental stations	1
8. Facilities and secretarial help	6
9. Federal Government Atomic Energy, TVA, Research Corp., Navy	1
10. Some leave arrangements	3
11. Expenses for material	2
12. No assistance	44
Total	69

Forty-four of the 69 institutions responding give absolutely no assistance to faculty members interested in research, whereas 25 give some form of assistance, ranging from expenses for material to the Carnegie Grants. The institutions which participate in Carnegie Grants are Fort Valley State College, Talladega College, Southern University, Tuskegee Institute, Morehouse College, and Atlanta University. These grants run up to \$5000 per year.

It would appear that institutions that fail to give any aid to research workers might encourage the efforts of these people by reducing their teaching load, but this case was found very discouraging. Of the 69 institutions responding to the questionnaire, only 16 or 23.2 percent reduce the load of teachers who are engaged in research, and, strange as it seems, 14 of the 16 institutions which reduce the teaching loads are institutions which are included in the 25 that are giving some form of aid to researchers.

### Conclusion

In brief it must be said that of the 69 institutions participating in this project 26 have research committees and 4 have directors of research. In these institutions are published two research bulletins; 10 use special sections of their catalogs or bulletins for faculty research work. Ten of these institutions have faculty members that have published 30 books within the last five years.

It seems unmistakably clear that a few Negro institutions are putting forth every possible effort to encourage research. This they are doing by offering such aid as the expense of material, secretarial help and facilities, small temporary grants, reasonable time off from teaching, reduction in teaching load and some travel funds. A few have even sought outside help of agencies such as General Education Board, The Carnegie Foundation and the Federal Government. In most of the institutions, however, people who do research must carry on their work with no help or consideration with regards to assistance or a reduction in their teaching load.

# INSTRUCTION AND PERSONALITY AS FACTORS IN STUDENT PERFORMANCE

BARRY T. JENSEN\*  
Personnel Research Section,  
AGO\*\*

NUMEROUS investigators have reported studies regarding the effect of instruction upon student performance and varying results have been noted. The writer found no differences among students taught by different instructors and students not attending class. The present study was an attempt to answer two questions: (1) in an independent study situation are students likely to acquire as much knowledge as those who attend class? And (2) what personality differences exist among students who are more or less successful when working independently?

The study discussed here is described in two parts as a matter of convenience. Part I deals with the effect of instruction; Part II deals with attempts to measure personality factors related to performance in school.

## Part I. The Effect of Instruction

Experiment 1: Method and results. — Since the core of the study hinged upon the writer's instruction during the second semester of the academic year 1950-51, it was necessary to determine if the sample was representative of his usual teaching. Accordingly, Experiment 1 was designed to test the hypothesis that there was no difference in effect of instruction from the first semester to the second.

A course in adolescent development was offered to Juniors, Seniors, and graduate students seeking to qualify for a teacher's certificate. Group I consisted of 74 students completing the course during the first semester; Group II consisted of all 56 students completing the course in a regular section during the second semester. The two groups were assumed to be equal initially.

\*Acknowledgement is made of the assistance of Dave Espin, graduate student at the University of Kentucky.

\*\*The study was completed at the University of Kentucky and was supported by a research grant from the University. It in no way reflects the opinions of the Department of the Army.

Procedure was similar for the two groups. Class meetings were held three days per week and activities consisted of lectures, movies, discussions and group conferences. Each student was given a topical study guide which contained suggestions for readings. Each topic listed was discussed in class. No definite reading assignments were given except that the students were informed that they could expect to be quizzed on the text references — about 300 pages. Text references were not given for all topics. Each student was required to submit reports on 900 pages of readings. Comparisons were based upon the student's performances on Part I of the final examination: "A Test of Knowledge of Fact and Principle in Human Development" by John E. Horrocks and M. A. Troyer.

There was no real difference in achievement of the two groups, as measured. The difference of .25 raw score points had a SE of .902.

Experiment 2: Method and Results. — This experiment was designed to test the hypothesis that there was no difference in the effect of attending the writer's classes and studying the same subject independently. Group A consisted of 42 students who were enrolled in and elected to attend classes for a course in adolescent development (These subjects were part of Group II described in Experiment 1).<sup>1</sup>

Group B consisted of 23 of the students electing the same course but who, when offered a choice, chose the section which would not meet in class and in which the students worked independently, as described below. Group B met once — the first day of the term. At that time the students received the study guide, a brief introduction to the subject (the same as given to other groups) and instructions to see the writer when they wished special help or desired to take a mid-term examination. They were told that they could expect the mid-terms to deal with material included in the study outline. No reports on reading were required of this group.

Comparison of Groups A and B was based upon performance on the final examination which had two parts: the test of knowledge of fact and principle, mentioned in connection with Experiment 1, and a portion of another of the Syracuse test series — "A Study of Barry Black." Table I contains information regarding Experiment 2. As can be noted, the two groups were similar in ability and in performance on the case-study test; Group A, which attended class, was superior on the test of knowledge of fact and principle.

## Part II. Personality and Performance

Personality tests and inventories were distributed to Groups A and B

1. Group A did not include all of Group II because data used in the second experiment could not be obtained for all subjects.

TABLE I

## COMPARISON OF TWO GROUPS: GENERAL ABILITY, FACTUAL KNOWLEDGE AND APPLICATION OF INFORMATION

Test	Mean	SD	Diff	t
<b>General Ability</b>				
Group A	.125	.088	.001	.045
Group B	.126	.088		
<b>Fact-Principle</b>				
Group A	54.29	2.32	4.83	5.188*
Group B	49.46	4.12		
<b>Barry Black</b>				
Group A	31.13	3.60	1.83	.932
Group B	30.30	3.34		

\*Significant at or beyond the .001 level.

at the meetings on the first day of the semester. Students were asked to return them within the week. The tests were distributed in packages; each test in a package carried the same number which also appeared on a slip of paper attached to the package. Each student wrote his name on the slip and gave it to the writer. This was done to give anonymity and yet permit follow-ups if tests were not returned soon. No papers were used if not returned within three weeks. Several students never did return the papers, even with several reminders.

Instruments used in this study were:

The California Test of Personality, Adult Level, Form A, which is an 180-item questionnaire that yields two part scores and a total score.

The Washburne Social Adjustment Inventory which yields several part scores and a total score.

An Index of Adjustment and Values by Bills, Vance and McLean. The "Index" yields two scores: (1) discrepancy between what one would like to be and what he thinks he is; and (2) a score indicating satisfaction with himself.

Education Opinionnaire, devised by the writer. It consisted of 70 statements with which the respondent indicated the extent of his agreement or disagreement. Some of the questions concerned school but many dealt with political and social problems — it was believed that answers to this instrument would indicate relative independence or maturity.

### Results

No total score and only one part score obtained by the administration of these personality measures discriminated between the group of persons choosing to attend class or to study independently. The group choosing independent study scored slightly but significantly higher on the "P" or purpose scale of the Washburne test but there was considerable overlap in the scores of the two groups with some persons in each group scoring at the extremes. Six persons enrolled in the independent study group (Group B) did not return questionnaires; if they had results might have been different inasmuch as these six students did not complete the course, either.

Table II contains information relative to correlations between various personality scores and achievement. Except for the measure of general ability (Kentucky Classification Battery — Part I) and 18 of the items of the Education Opinionnaire, none of the tests used was significantly correlated with success under the independent study procedure. (General ability and Education Opinionnaire scores for the independent study group was .387, significant at the .05 level.)

The 18 items of the Education Opinionnaire were retained because they discriminated between the more and the less successful persons in Group B (those choosing not to attend class). Items which showed a

TABLE II

## CORRELATIONS BETWEEN VARIOUS PERSONALITY SCORES AND SUCCESS UNDER CONDITIONS OF INDEPENDENT STUDY\*

Personality-Measure	Fact-Principle	Study of Barry Black
Washburne (total)	-.185	.050
Index (Satisfaction)	.009	-.177
Index (Discrepancy)	.271	.188
California (total)	.139	-.288
General Ability	.546**	.526**
Educational Opinionnaire 18 items	.721**	.608**
Educational Opinionnaire 18 items, General ability partialled out	.676**	.540**

\*Correlation coefficient of .219 was obtained between the two criterion measures.

\*\*Significant at or beyond the .01 level.

score difference of one-half point or more between high or low halves were retained and a new scoring key was based upon the answers given by the more successful group.

The papers were rescored. The difference between means of the high and low halves of Group B (on the basis of factual knowledge) was significant at the .01 level. As indicated in Table II, scores on these eighteen items correlated .721 with scores on the test of knowledge; with general ability partialled out the correlation coefficient was reduced to .676. The correlation between the eighteen items of the Education Opinionnaire and the Study of Barry Black was .608 which was reduced to .540 when general ability was partialled out.

Reliability estimates were based on odd-even scores obtained by different groups of students. One coefficient was negative and the other was not significantly different from zero. The assumption of parallelism between halves of the test was not justified.

### Summary and Conclusions

Comparisons were made of the performance of students electing the writer's course in adolescent development during either first or second semesters of 1950-51. Comparisons were also made of the performance of the second semester group and students electing to study the subject independently. A number of personality instruments were given to students at the beginning of the second semester. There was no difference in achievement of the groups attending class during either semester. It is concluded that the instruction during Experiment 2 was a good sample of the writer's teaching.

There was no difference in achievement by the class attenders and the independent study group when the criterion was performance on a test of application of information (Study of Barry Black) but when the criterion was knowledge of facts and principles the group attending class scored significantly higher than did the non-attenders.

On the basis of total personality test scores it was not possible to distinguish between the students electing to attend class and those electing to study independently. Neither was it possible to discriminate, on the basis of personality test scores, between the more and the less successful students under conditions of independent study.

For the independent study group score on the Education Opinionnaire was significantly correlated with success on each of two uncorrelated criterion measures, but the same relationship was not found for the group attending class. It is concluded that the Education Opinionnaire sampled something important to success per se under independent study conditions. That this something is not general ability is shown by the correlation shown after general ability was partialled out. The selected questions on the opinionnaire were related to self-reliance, attitude

toward school, and motivation.

Since reliability of the Education Opinionnaire was not established, care must be used in making generalizations. However, the relationships observed are suggestive that further study in this direction might be of value.

On the basis of two studies it is concluded that students not attending class can achieve as much as those attending class regularly if the course is not rigidly outlined and students have to work somewhat independently; those attending class might achieve more when the criterion is acquisition of knowledge about the course content.

#### REFERENCES

- R. E. Bills, Edgar L. Vance, and Orison S. McLean, "An Index of Adjustment and Values," Journal of Consulting Psychology, XV (1951), 257-261.
- Barry T. Jensen, "A Comparison of Student Achievement Under Conditions of Class Attendance and Non-Attendance," College and University, XXVI (1951), 399-404.

# A COMPARISON OF SUPERVISED CORRESPONDENCE STUDY PUPILS AND CLASSROOM PUPILS IN ACHIEVEMENT IN SCHOOL SUBJECTS

G. B. CHILDS

University of Nebraska  
Lincoln, Nebraska

## Introduction

THE LAST twenty-five years have witnessed the development of a new educational instrument in America. This instrument is Supervised Correspondence Study. Developed as a device for enriching the curricular offerings of small high schools and as a means of bringing education to pupils isolated from regular schools by factors of distance or physical handicaps, this method of instruction has grown until today many thousands of young people receive all or part of their secondary schooling in this way.

Despite the position of importance which Supervised Correspondence Study has come to occupy there is little research concerning this means of instruction. Most of the research which does exist is concerned with details of procedure. Very few studies attempt to evaluate the soundness of Supervised Correspondence Study as an educational device and those studies which do present evidence on this point were conducted before the time when high school courses were prepared according to carefully devised techniques, are limited in regard to the number of pupils involved, or deal with only a relatively small number of school subjects or with a limited subject matter area. Moreover, in few of the studies were control groups established.

## Problem

Recently the writer completed a study<sup>1</sup> which was more detailed than previous studies in this field. Specifically stated, the problem of the investigation was: How does the achievement of pupils who take courses by Supervised Correspondence Study compare with the achievement of pupils who receive instruction in similar courses in regular classroom situations?

1. G.B. Childs, A Comparison of Supervised Correspondence Study Pupils and Classroom Pupils in Achievement in School Subjects. Unpublished Ph.D. dissertation, Teachers College, University of Nebraska, 1936.

Procedure

In conducting the study, fourteen different subjects from five different subject matter fields were used. To secure a control group of classroom pupils for each subject, five Nebraska schools in which the subject was offered were selected for inclusion in the study. Selection of schools was largely random except that small schools were used whenever possible. This practice was followed because a very large majority of correspondence pupils come from such schools and it was desired to compare the correspondence pupils with other pupils in similar school situations.

The correspondence group consisted of those pupils who completed courses in the fourteen subjects of the study between March 1, 1948 and April 1, 1949.

The subjects included in the study were:

First-year algebra	Physics
Third-semester algebra	Chemistry
Plane geometry	American history
Solid geometry	World history
Trigonometry	First-year bookkeeping
First-year Latin	First-year typewriting
First-year Spanish	First-year shorthand (Gregg)

The Cooperative tests prepared under the direction of the American Council on Education were used to test achievement in all subjects except those in the area of commercial arts; i. e., typewriting, shorthand, and bookkeeping. In the commercial subjects the tests prepared by the United States Armed Forces Institute were used.

Since the study called for a comparison of the achievement of two groups of pupils, it was essential that information be made available which would enable pupils of similar ability to be compared. To this end the Henmon-Nelson Test of Mental Ability for Grades 7-12, Form C, was given to each pupil included in the study.

To secure further background material the USAFI Tests of General Educational Development (hereafter referred to as GED tests) were used to test pupils in those subject areas for which such tests are available. Specifically, these tests were employed to secure information concerning the background which correspondence and classroom pupils in beginning algebra, plane geometry, physics, chemistry, American history, and world history brought to the study of these subjects.

The GED tests were administered to classroom pupils near the beginning of the semester in which they were enrolled in the subjects in connection with which this test was used. This test was administered to all correspondence pupils in the same subjects shortly after the

time of their enrollment. The achievement tests and the intelligence test were administered to all pupils at the time they completed work in the courses they were carrying.

### Analysis of data

The basic statistical procedure<sup>2</sup> used in this study is an adaptation of the method of analysis of variance which permits the utilization of the full distribution of two groups without rejecting cases to secure matching.

In applying this method the two groups are subdivided into smaller classes and subclasses thus providing a number of categories which are relatively homogeneous with respect to the background trait or traits which are employed for matching. In this study the primary classes were formed on the basis of intelligence, and subclasses were based on chronological age. These background factors were used in applying this statistical technique to test results obtained in first-year algebra, third-semester algebra, solid geometry, trigonometry, Latin, typewriting, shorthand, American history, and world history. A separate analysis was made using intelligence and scores on the Tests of General Educational Development as background factors, for first-year algebra, American history, and world history.

In using this statistical process the mean variance between groups and the mean variance within groups are determined and the value of F is found by dividing the larger mean variance by the smaller mean variance. It is then possible to determine the significance of this obtained F by entering Snedecor's Table<sup>3</sup> for F with the appropriate level of confidence desired. The null hypothesis tested is that there is no difference between the achievement scores of correspondence and classroom pupils. In this study values of F at both the 1 per cent and 5 per cent levels of confidence will be reported.

The results of the analysis made through use of the above procedures are shown in Table I.

In several of the one-year subjects included in the study, the number of test scores obtained for correspondence students who had taken both semesters of work by correspondence was too small to permit them to be organized into logical classes and subclasses. It was, therefore, necessary to employ some other method than that reported above. The method used was to match each correspondence student with all classroom students who possessed approximately the same

2. Palmer O. Johnson and J. Neyman, "Tests of Certain Linear Hypothesis and Their Application to Some Educational Problems," Statistical Research Memoirs, I (June, 1936) pp. 57-93.

3. George W. Snedecor, Statistical Methods (Ames: The Iowa State College Press, 1946), pp. 222-225.

TABLE I

RELATIONSHIP BETWEEN ACHIEVEMENT OF CORRESPONDENCE PUPILS AND CLASSROOM PUPILS IN SELECTED SUBJECTS AS DETERMINED BY AN ADAP-TATION OF THE METHOD OF ANALYSIS OF VARIANCE

American History	IQ and age					
Between Groups		1	346.33	4.42*	6.90	3.94
Within Groups		97	78.43			
American History	IQ and GED					
Between Groups		1	87.17	1.49*	6.90	3.94
Within Groups		100	58.60			
World History	IQ and age					
Between Groups		1	506.23	5.98*	6.90	3.94
Within Groups		98	84.70			
World History	IQ and GED					
Between Groups		1	72.76			
Within Groups		98	79.98	1.10*	6334	253
Elementary Algebra	IQ and age					
Between Groups		1	389.41	6.26*	7.13	4.02
Within Groups		54	62.16			
Elementary Algebra	IQ and GED					
Between Groups		1	201.32	5.99*	7.16	4.03
Within Groups		51	33.58			
Intermediate Algebra	IQ and age					
Between Groups		1	305.86	8.20*	7.42	4.12
Within Groups		35	37.32			
Solid Geometry	IQ and age					
Between Groups		1	30.84			
Within Groups		51	70.82	2.29**	6303	252
Trigonometry	IQ and age					
Between Groups		1	113.84	3.32**	7.44	4.13
Within Groups		34	34.30			
Latin	IQ and age					
Between Groups		1	417.67	12.96*	7.21	4.05
Within Groups		46	32.22			
Typewriting	Speed	IQ and age				
Between Groups		1	14219.03	96.23*	6.80	3.91
Within Groups		164	147.76			
Knowledge and Skills	IQ and age					
Between Groups		1	8538.84	50.51*	6.80	3.91
Within Groups		163	169.05			
Shorthand	Vocabulary	IQ and age				
Between Groups		1	1193.63	1.88*	7.27	4.07
Within Groups		42	635.96			
Dictation and Transcription	60 wpm	IQ and age				
Between Groups		1	2.86			
Within Groups		41	76.47	25.74*	6286	251
70 wpm	IQ and age					
Between Groups		1	13.45			
Within Groups		41	155.99	11.60**	6286	251

\* The difference in achievement of the groups favors the correspondence group.

\*\*The difference in achievement of the groups favors the classroom group.

background factors. If the bases for matching were IQ and age, then the IQ and the age of a correspondence student were found and all classroom students in the same subject of the same age and with approximately the same IQ were matched against him. The achievement scores of the classroom pupils matched with the correspondence pupils were averaged to secure a single score which represented the score of a hypothetical student matched against the correspondence student. The average used was the mean.

After the matching process had been completed, the mean of the achievement scores of the correspondence pupils in a particular subject was found and compared with the mean of the matched scores to determine the significance of existing differences. This was done by means of the *t* test of significance. Again the null hypothesis tested is that no difference exists between the two groups.

The test results for plane geometry, Spanish, physics, chemistry, and bookkeeping were analyzed by this method when IQ and chronological age were used as matching factors. In plane geometry, physics, and chemistry, separate comparisons were made using IQ and GED scores as the bases for matching.

The results of the analyses made by use of the *t* test for significance are reported in Table II.

#### Report of findings

When matching factors were IQ and chronological age, results obtained by use of the method of analysis of variance show that:

1. The mean scores of correspondence pupils on the achievement tests exceeded the mean scores of classroom pupils by an amount which was significant at either the 1 per cent or 5 per cent level of confidence in American history, world history, first-year algebra, third-semester algebra, Latin, typewriting speed, and typewriting knowledge and skills.

2. There was no significant difference between the mean scores of classroom pupils in shorthand, solid geometry, and trigonometry.

When matching factors were IQ and GED scores, results obtained by the method of analysis of variance show that:

1. The mean score of correspondence pupils exceeded the mean score of classroom pupils by an amount which was significant at the 5 per cent level of confidence in first-year algebra.

2. The mean scores of correspondence pupils exceeded the mean scores of classroom pupils by an amount which was not significant in American history and world history.

TABLE II

RELATIONSHIP BETWEEN ACHIEVEMENT OF CORRESPONDENCE PUPILS AND  
CLASSROOM PUPILS IN SELECTED SUBJECTS AS DETERMINED BY THE  $t$  TEST  
OF SIGNIFICANCE

Subject	Background Factors	Degrees of Freedom	Mean Difference*	Obtained t	t for 1%	t for 5%
Plane Geometry	IQ and Age	7	10.99	2.754	3.499	2.365
Plane Geometry	IQ and GED	6	8.17	1.870	3.707	2.447
Spanish	IQ and Age	6	16.97	3.746	3.707	2.447
Physics	IQ and Age	4	5.27	1.959	4.604	2.776
Physics	IQ and GED	4	3.87	1.142	4.604	2.776
Chemistry	IQ and Age	4	1.10	.339	4.604	2.776
Chemistry	IQ and GED	4	1.27	.606	4.604	2.776
Bookeeping	IQ and Age	5	2.32	.459	4.032	2.571

\*The mean difference in all comparisons favors the correspondence over the classroom group.

Results obtained by use of the t test for significance show that for pupils matched on the basis of IQ and chronological age:

1. The mean scores of the correspondence pupils exceeded the mean scores of the classroom pupils by an amount which was significant at either the 1 per cent or 5 per cent level of confidence in plane geometry and Spanish.

2. The mean scores of correspondence pupils exceeded the mean scores of classroom pupils by an amount which was not significant in physics, chemistry, and bookkeeping.

Results obtained by use of the t test for significance show that for pupils matched on the basis of IQ and GED scores:

1. The mean scores of correspondence pupils exceeded the mean scores of classroom pupils by an amount which was not significant in plane geometry, physics, and chemistry.

### Conclusions

1. Since the achievement of correspondence pupils tested for this study tended to exceed the achievement of classroom pupils against whom they were matched, pupils may be registered for instruction by correspondence study with a high degree of assurance that their mastery of the subject matter of the courses taken will be satisfactory.

2. There is a difference in the extent to which different correspondence courses are successful, when success is measured by the margin of difference between the performance of correspondence and classroom pupils. The range is from a very marked difference in favor of the correspondence group in typewriting to a difference which approached significance in favor of the classroom group in trigonometry.

3. The achievement of correspondence pupils tended to exceed that of classroom pupils by a greater margin when IQ and chronological age were used as matching factors than when IQ and GED standard scores were used for this purpose. If the GED test scores can be taken as a valid measure of the ability of students in the fields in which the GED tests were given, this result indicates that pupils who complete courses by correspondence are, by and large, more capable students than pupils of the same IQ and age in regular classrooms. The probability that pupils who complete correspondence courses are a rather select group is further attested by the fact that, aside from typewriting, few pupils who completed correspondence courses and who were tested for this study had low intelligence test scores.

# SCHOLASTIC BACKGROUND AS RELATED TO SUCCESS IN COLLEGE PHYSICS

SAM ADAMS, McNeese State College,  
Lake Charles, Louisiana

H. L. GARRETT, Louisiana State University, University, Louisiana

FOR MANY YEARS certain educators have felt that somehow there should be a means of predicting the probable success of a student after college entry. A great deal of work has been done, therefore, in attempting to find interrelationships between a student's scholastic background and his record of achievement in college.

Numerous problems have arisen to plague such investigators. Some of these have been: (1) How can achievement be measured? (2) How can one compensate for variations of marking systems between schools, or between individual teachers in the same school? And (3) how can personal factors — ambition, persistence and similar qualities — be considered?

The result of many such studies has been the setting off of a new series of pot-shots between high school and college teachers. The claim by some college teachers that high schools have omitted subject matter from the curriculum is frequently countered by the remark that things would be quite different if the college teachers had the slightest idea as to how to teach.

## Problem

Many investigations regarding factors which contribute toward success in college have dealt with large areas, such as overall averages or point-hour ratios. Somewhat fewer studies have been made regarding success in particular courses. In very few cases has work been done concerning success in the course considered here — college physics.

Specifically, the problem was to find what relationships, if any, existed between achievement in college physics and achievement in various background items, including: high school courses, college entrance tests and first year college mathematics.

## Data

This investigation dealt with 877 students who took a beginning phys-

ics course at Louisiana State University. These students were selected on the following bases: (1) They began a physics course in the fall semester of 1947, 1948 or 1949; (2) they completed the second term of the course in the spring semester of the same school year in which they began the study of physics; and (3) their high school records were available.

Individual data cards were used, and the following data were taken from the high school transcripts: marks in high school physics, chemistry, algebra II and senior English; averages in high school science, English and mathematics; and percentile rank in high school class. From the university records the marks in first year college entrance tests were recorded: the Purdue Placement Test in English, the Cooperative Test of General Proficiency in the Field of Mathematics, the American Council on Education Psychological Examination (Ranks on the total test and rank on the "quantitative thinking" part of this test were recorded separately), the Iowa Silent Reading Test, and a locally-prepared chemistry test. Composite ranks on the entire test series were calculated for those with complete test records.

#### Treatment of Data

A large majority of high school marks, as well as those given at the university, followed the sequence A-B-C-D-F. These marks were assigned numerical values (A=4 to F=0); for a two-semester course, such as most of these were, the two semester mark equivalents were added, giving year marks ranging from eight (for two A marks) to zero. Course marks and averages in the various fields were reduced to year marks for statistical treatment.

Percentile ranks were used as indicators of achievement on the various entrance tests. Ranks in high school class were reduced to a percentile basis.

A common method for studying relationships between two variables is by means of the coefficient of correlation,  $r$ . Several types of procedures can be employed; however, for the purposes of this study the Pearson product-moment type of correlation was used. In each case the probable error (PER) was calculated, this serving as one method of ascertaining whether or not a particular coefficient was meaningful. A convenient rule-of-thumb is that a relationship probably exists if the coefficient ( $r$ ) is as much as four times as large as the probable error of the coefficient.

These data were reduced to a simple code, so that IBM equipment could be used in setting up the various scatter diagrams used in calculating coefficients of correlation.

Results

Table I gives a summary of the results of this investigation.

Many of these coefficients were relatively low. However, only in the last case — that relating success in college physics with rank on the reading test — was it so low as to be of doubtful significance.

The highest coefficient was that relating college physics and college mathematics. There are at least two factors which might explain this: (1) Many college physics teachers place great emphasis on problem-solving, and (2) the mathematics course was taken just before the physics, so that the element of recency probably made a difference.

Subject articulation between high school and college physics ( $r = .324$ ) was relatively poor. Some college physics teachers have argued that students would be better prepared for college physics if high school teachers would stop teaching the course from the college-preparatory point of view. This apparently anomalous position they explain thus: Present courses put too much emphasis on techniques and problem work, both of which are readily forgotten. If the emphasis were on basic principles, more retention would be in evidence. However, high school teachers are quick to point out the abstract presentation commonly used in college physics classes, and many can always illustrate their argument by describing some notably poor college teaching.

The low correlation between college physics achievement and high school science average ( $r = .204$ ) is probably to be expected. Many students had a two-year science program in high school, this consisting of general science and biology. Obviously there is not a great deal in such a program that would carry over into college physics. Too, several years intervened between the time the student finished his high school science work and the time he began college physics.

During recent years there appears to have been a trend toward wider use of tests and less use of high school records in evaluating the college potentialities of students. In part this might be attributed to greater facility in compiling and using test data. It appears doubtful, however, if test records based on the work of a few hours could tell as much about a student as would a high school record based on a four-year period. These data appear to bear out this point, since six criteria — first year college mathematics, high school physics, rank in class, algebra II, high school mathematics average and high school chemistry — all have higher coefficients of correlation with achievement in college physics than does the highest of the entrance tests.

Conclusions

On the basis of the findings of this study, some general conclusions

TABLE I

## CORRELATION BETWEEN COLLEGE PHYSICS MARKS AND VARIOUS PHASES OF SCHOLASTIC BACKGROUND

Items From Scholastic Background	Number of cases	Percent of Total Group	r	PER
First year college mathematics	340	39	.435	.035
High school physics	367	42	.324	.032
Rank in high school class	518	59	.306	.027
Algebra II	506	58	.290	.028
High school mathematics average	655	75	.279	.041
High school chemistry	498	57	.263	.028
Mathematics entrance test	351	40	.258	.050
High school senior English	622	71	.247	.037
High school English average	659	75	.241	.025
Composite rank on entrance tests	320	36	.214	.036
High school science average	607	69	.204	.026
Chemistry entrance test	360	41	.187	.034
English entrance test	411	47	.154	.032
Psychological test (total)	422	48	.130	.032
Psychological test (quantitative)	418	48	.121	.030
Reading test	423	48	.077	.032

appear to be justifiable.

1. Articulation between college physics and various types of high school work was relatively poor.
2. Despite this, high school records appeared to tell more about probable success in college physics than did entrance test ranks.
3. A relatively high relationship appeared to exist between achievement in college physics and achievement in first year college mathematics.

Despite the fact that much work has been done regarding various factors as related to success in college, there is little reason to believe that final answers are at hand. In short, any system of student personnel work which is based upon reducing students to the status of index numbers or correction factors is probably doomed to failure.

# SCIENCE INFORMATION AND ATTITUDES POSSESSED BY CALIFORNIA ELEMENTARY SCHOOL PUPILS

STANLEY B. BROWN  
University of Colorado  
Boulder, Colorado

IT HAS BEEN generally recognized that very little data is available on the achievement of elementary school pupils in science information and science attitudes. To attack that problem of insufficient data, a study was undertaken to analyze the science information and attitudes possessed by California elementary school pupils. The children were selected in terms of geographical location, 5th and 8th grade levels, and sex.

## Procedure

Fundamental criteria were developed to serve as a basis for the analysis and the following question-type statements were used in the study:

1. Is a functional understanding of facts, principles, and concepts of science developed?
2. Are the areas of physical and biological science covered adequately?
3. Do the pupils realize and appreciate the interdependency of living things and their relationships to human institutions?
4. Is an appreciation of an orderly universe operating under natural laws developed?
5. Are the pupils sensitized to the impact of scientific development on civilization?
6. Is the gap between scientific technology and curriculum content reduced?
7. Is a scientific vocabulary developed?
8. Are pupils acquainted with the sources of scientific knowledge?
9. Is the scientific attitude and method of critical thinking developed?
10. Is an appreciation of constantly improving living standards through scientific achievements developed?

\* Abstract of a dissertation submitted to the Graduate School of Education, Stanford University, California, in partial fulfillment of the requirements for the degree of Doctor of Education (1951). Dissertation prepared under the direction of Assistant Professor B. Frank Gillette, Professor Quinn McNemar and Professor Paul Hanna.

11. Is an understanding of natural resources including their relationship to human institutions and the necessity for their conservation developed?
12. Is a functional understanding of health information and desirable hygienic habits developed?

Science information and attitude tests were prepared and administered in Spring of 1951 to 2901 pupils in forty-one selected California elementary and junior high schools representing fourteen urban, nine suburban, and eighteen communities. Areas included in the instruments were biological sciences (including health information), physical sciences, conservation of natural resources, and consumer education.

Validity of the instruments was established through the pooled judgment of competent observers from the fields of both science and science education. Reliability was determined by the split-halves technique and found to be .90 for the science information and .73 for the science attitude test.

### Findings

With reference to geographical location, there was found to be no appreciable difference between the urban and suburban pupil's level of achievement on the science information test. The differences between the urban versus the rural and suburban versus the rural pupils respectively indicated a slight gain in favor of the rural pupils. These differences were found to be reliable as shown by the critical ratio.

In comparing the science attitude test results in terms of geographical location, the findings revealed that the rural pupils possess more favorable attitudes than the pupils of suburban and urban areas. The urban and suburban groups were found to possess somewhat similar attitudes and of a lesser rank than the rural groups.

On the science information test endorsed by validating experts as representative of what should be taught in the elementary school science program, 8th grade pupils answered somewhat more than three-fourths (82 percent) of the items correctly, whereas the 5th grade pupils answered two-thirds (66 percent) of the items correctly. The science attitude test findings revealed that the correct or favorable responses by 5th grade pupils were 58 percent of the items and the 8th grade pupils 63 percent. In other words, more than half of the pupils in the study gave an undesirable response to the science attitude test. These data indicate that 8th grade pupils, irrespective of sex, were found to make significantly better scores on the science information test than did the 5th grade pupils, and although the increment between the 5th and 8th grade pupils' attainment of science attitudes was somewhat less than for the science information test section, the improve-

ment that does exist was statistically significant as determined by the critical ratio.

The boys in both the 5th and 8th grade groups obtained somewhat higher scores on the science information test than did the girls. The differences were found to be statistically reliable.

The 5th and 8th grade girls achieved slightly higher scores than did the boys on the science attitude test; however, the differences were unreliable from a statistical viewpoint.

The average correlation between the science information and attitude tests was found to be .47, indicating a moderate relationship between the two tests. The range of correlation for the several groups extended from .34 for the 5th grade suburban girls to .67 for the 8th grade suburban girls.

### Conclusions and Recommendations

The following conclusions and recommendations were based upon the findings of this study as they were related to the judgments of prominent science education authorities:

1. Since 5th and 8th grade boys scored somewhat higher on the science information test than the girls, this differential suggests that the science program is set up to meet the needs of the boys' interests and presumably fails to comply with feminine applications.
2. The findings on the science attitude test revealed that the girls achieved slightly higher scores than the boys; however, this was statistically shown to be an unreliable differential. No conclusions can be drawn from this particular phase of the study, thus substantiating the belief that a need for further investigation with a more comprehensive science attitude instrument would help clarify this limitation.
3. Significant differences were found between the attained mean scores of the 5th and 8th grade pupils when compared on the science information test. This leads to the conclusion that the kind of science information is being expanded or built up during these three formative years.
4. In comparing the percentages of favorable responses to the attitude test items by the 5th and 8th grade pupils, only a slight gain was apparent. These data indicate that the development of desirable science attitudes is achieving only limited success in California elementary schools. These findings also substantiate the statement that one cannot assume the acquisition of desirable attitudes as always accompanying the learning of factual information. Since the findings also revealed that rural pupils possess superior science attitudes, it seems apparent that urban and suburban schools should

improve their methods of teaching proper science attitudes. Educators agree that the most effective way of developing desirable attitudes is by teaching directly for them rather than expecting their realization through transfer resulting from the amount of science information mastered.

5. Adequate evaluation instruments should be developed and made available to educators. Investigations and evaluative instruments in the field of elementary science education are quite limited. Because of the vital importance of the program, more research is needed to develop curricular sequence, techniques of teaching, audio-visual materials, teaching aids, and valid and reliable evaluative instruments in the field.
6. Since all elementary teachers have responsibilities in science education they should be encouraged to take science education courses and experiences both at the pre-service and in-service levels.

The vital importance of scientific enlightenment for the average individual is becoming more obvious in this modern world. The attainment of a functional understanding of science principles, facts, and concepts with particular emphasis on the development of desirable science attitudes constitutes a real challenge if tomorrow's citizens are to live effectively in one world!

# CHECK-LIST FOR DOCTORAL RESEARCH

JOHN W. KIDD  
Michigan State College  
East Lansing, Michigan

IN ATTEMPTING to help certain graduate students develop more desirable perspective in relation to contemplated research and at the same time to provide them with systematic safeguards as they proceed, the writer has devised and utilized the following list of things to consider. While most of the points covered are applicable to any research, the total list is particularly aimed at the practical situation faced by the typical doctoral candidate.

## Preliminary

### 1. Originality.

Perhaps the most useful criterion to apply is that of "emergent knowledge" rather than the sometimes exaggerated notion of discovery, invention, or creativity.

### 2. Feasibility.

a. Sufficiently narrowed down, pin-pointed.

b. Accessibility of data.

c. Maximum advantage taken of researcher's unique position; may provide relatively easy job; researcher may be in position to probe where others do not have access; use language facility, experience, residence, past and present employment.

d. Reasonableness of size of undertaking.

e. Favorable time factor.

f. Favorable cost factor.

g. Special techniques and abilities needed and available such as: statistics, sociometry, sampling, questionnaire.

h. Information necessary to the utilization of special techniques, as in the case of questionnaire: needed returns, probable returns, avoiding bias, establishing validity and reliability, analyzing returns by IBM cards, etc.

### 3. Defensibility.

a. Reasonableness of hypothesis(es).

b. Current interest.

c. Findings likely to go beyond readily available knowledge.

**4. Publishability.**

Likely to see print in whole or part.

**5. Applicability.**

Contribution to knowledge; effect on candidate and total of human knowledge; possible implications.

**6. Respectability.**

Professionally, intellectually, scientifically; not in term paper category.

**7. Acceptability.**

"Selling" the idea to major professor, committee, graduate school, institution; know the kind of things previously accepted and, if possible, rejected by each; detect interests of key persons.

**Planning****1. Validity.**

Does the purported job; inclusiveness and exclusiveness.

**2. Reliability.**

Does the job well; internal and external consistency; avoiding the influence of weak or inaccurate observations; the nature of fact.

If necessary, be familiar with techniques of

- Pilot study.
- Retesting.
- Interview — structured and unstructured.
- Split-halves technique.
- Utilization of outside criteria.

**Execution****1. Communicability.**

Beware of unwarranted assumptions as bases of presentation; have understanding with committee as to level of understanding to be assumed on the part of the reader.

**2. Objectivity.**

Danger of pre-judgments, habitual attitudes, loaded words and concepts such as "contribution" and "influence," attempt neither to prove nor disprove but to reveal reality.

3. Specificity.
  - a. Delimitation of scope.
  - b. Continuing check on relatedness of material.
  - c. Restricting generalizations to those warranted by the evidence — danger of over-generalizing as opposed to justifiable extrapolation and statements of probability.
4. Propriety.  
In good usage, words, grammar, composition, consistency, pagination, divisions and subdivisions, titles, tables and figures, margins, bibliographical materials and footnoting, etc.

#### General

1. Get publishing habit. Know professional journals and their policies. Many term papers are publishable. Publications a vital key to professional advancement.
2. Use systematic approach. Card file or other technique in early reading and experience in relation to possible research.
3. Need for wide acquaintance in research area. Researcher in role of expert; should know more about particular research problem than anyone else.

## Research News and Communications

Lawrence P. Blum, University of Wisconsin Extension, Milwaukee; Jacob O. Bach,  
Southern Illinois University; Lester Sands, Santa Barbara College

Study student personnel work as a factor in the accreditation of institutions of higher education. This project reports an exploratory study of the evaluation of student personnel work by those accrediting agencies which include this phase of college programs in their investigations. Four things are attempted. First, there is presented a review of historical development of standards and criteria used by accrediting groups in evaluating student personnel programs. Secondly, a critical examination is made of the present standards and the procedures used in applying them. This phase of the study led to the conclusion that (1) with the exception of those of the Middle States Association of Colleges and Secondary Schools, the accrediting standards for student personnel work at present call for largely quantitative measurement rather than qualitative evaluation of institutional programs; (2) that because the accrediting agencies set forth minimum essentials in their standards, the tendency probably is, if not to discourage, at least not to encourage, the development of new and unusual kinds of student personnel programs; and (3) that the methods by which the standards and criteria are applied frequently tend to be less democratic than might be thought desirable.

The third phase of the study reports investigations of the amount and kind of participation by student personnel workers in the activities of accrediting agencies. Of eight hundred and fifty-five workers replying to a short questionnaire, only fifty had ever participated, and of these, only eight had done so in capacities which indicated that they might have been able to influence significantly the standards and procedures used by some accrediting agency. A second, longer questionnaire investigated the familiarity of student personnel workers with the work of accrediting agencies, their opinions of the significance to student personnel work of the agencies' activities, and of the adequacy of standards and procedures. Suggestions for improvement were also solicited. The results indicate that the agencies have not been very successful in publishing their activities among, or in "selling" themselves to student personnel workers, and that many such workers hold quite negative opinions of accreditation in general, and the practices of some specific agencies in particular.

The fourth section consists of proposals which the writer feels might serve to make the work of the agencies more meaningful and more helpful to the institutions served. In brief, these are (1) that the evaluation of student personnel programs should be in terms of their effect on individual students; (2) that the accrediting agency's role in such a situation should be to evaluate the institution's research into its own student personnel needs in the light of its objectives, and into the degree to which the student personnel program contributes to the production of desirable educational outcomes as defined by the institution; and (3) that the agencies might make more significant contributions than at present to the improve-

ment of student personnel programs by supplying consultants and technicians to assist local student personnel workers to plan and carry out meaningful research programs.

Additional information concerning the study and its outcomes can be secured from James F. Penney, Teachers College, Columbia University.

Study desirable results and accident rates in trampoline instruction. An experimental study comparing changes in certain desirable physical abilities occurring in trampoline classes with changes in these abilities occurring in classes in soccer, hockey, speedball, and dance is being made. The study is being conducted by the comparative group method using seven classes of freshman women. Tests were given at the beginning of the term. These will be repeated at the end of the term and comparisons of the changes occurring will be made.

A large majority of trampoline instructors contacted claim that their students achieve greater increases in body control, agility, endurance and strength of muscles of the abdomen, shoulder girdle, back and legs than they would in other physical education activities. This study is an investigation of these claims. A survey of trampoline accident experience has been completed. This indicates that, when properly supervised, trampolining is no more dangerous than other physical activities. Since trampoline activity has a strong appeal for most students, if the claims of its benefits are well founded, the advisability of using it in the educational program would be strengthened.

Additional information concerning these studies can be secured from Helen Zimmerman, Southern Illinois University, Carbondale, Illinois.

AERA Fellowships in educational measurement. Arthur E. Traxler, chairman of the AERA Fellowship Award Committee, has announced that two fellowships in educational measurement have been awarded for the academic year 1953-1954. Because two of the applicants were so outstanding, the Fellowship Award Committee recommended that the World Book Company which furnishes the stipends award two fellowships for this year. The recommendation for this exceptional action was generously approved, for this year only.

The two recipients of the fellowship awards are Mr. Roman Stephen Gawkoski and ~~Mr.~~ David Brom Orr. Mr. Gawkoski holds the B.S. and M.A. degree from St. Louis University. Mr. Orr holds an A.B. and B.S. degree from Wittenberg College and has done graduate work at Teachers College, Columbia University.

What research says to the teacher. The American Educational Research Association and the NEA Department of Classroom Teachers, cooperating through the NEA Research Division, have begun a series of pamphlets interpreting the findings of research to the teacher. The first pamphlet - Teaching Reading by Arthur I. Gates - was published in June of this year. The price for a single copy is twenty-five cents. Lower prices prevail for quantity orders.

# JOURNAL *of* EDUCATIONAL RESEARCH

---

Volume XLVII

April, 1954

Number 8

---

## THE RELATIVE VALUE OF RECALL AND RECOGNITION TECHNIQUES FOR MEASURING PRECISE KNOWLEDGE OF WORD MEANING--NOUNS, VERBS, ADJECTIVES

DELPHA HURLBURT  
Pacific University  
Forest Grove, Oregon

PEOPLE CAN understand a great many more words while reading or listening than they are able to remember and recall quickly for use in speaking and writing. With words like nostalgia, sagacious, connoisseur and cognizant the average adult will be able to recognize and comprehend their general meaning when he hears them in a speech or reads them in stories or poems. Yet in his conversation or in his writing of letters the average adult practically never uses such words. Also some people are able to learn more words from their reading and listening experiences than other people. Furthermore, many people, who do enlarge their vocabulary from their reading and listening activities, are also capable of adapting and applying those words in their speaking and writing. Nevertheless, over and above such individual differences in the ability to understand and use words, vocabulary power or ability seems to vary with the type of mental and language activity being pursued. Hence, it appears that each person has an active or readily available vocabulary composed of the words he actually uses in ordinary communication and a passive or latent and potential vocabulary composed of words which he understands in reading but does not use in his original language expression.

### Purpose

The purpose of this study is to compare the active or writing and speaking vocabulary with the latent or reading vocabulary of high school students.

### Description of Study

Construction of testing instruments. — Separate tests of 115 items each were written for nouns, verbs, adjectives. Each test was prepared in two forms with identical items. A recall-completion form was designed to test active vocabulary or facility in summoning words for writing and speaking. A recognition-multiple choice form was designed to test an individual student's latent vocabulary or ability and associate meaning of precise words.

Words were selected by a random sampling of Webster's Synonym Dictionary and Mawson's Roget's Thesaurus in Dictionary Form for each part of speech. Phrase definitions for words were written in strict accordance with dictionary definitions and arranged in sections according to synonym categories. All three tests were equated for style of writing, type of item, reading difficulty, similarity of directions, mechanical make-up and format, and testing time.

Administration of vocabulary tests. — All vocabulary tests were administered in the regularly scheduled English-class periods by the regular English teachers. The recall-completion form of each test was always given ahead of the recognition-multiple choice form of the same test. Order of given tests was rotated so that every test would be given approximately equal number of times at each position in the test series. Each form of every test was given at a single sitting. Testing time for each test form was approximately 41 minutes — a full class period. In the majority of instances, a single test was given each week over a period of three weeks and not more than three days ever elapsed between giving of the recall-completion form and the recognition-multiple choice form of the same test.

Subjects. — One hundred ninety-two students from Grade IX and two hundred ten students from Grade XI at Revere High School were used for this experimental study. The mean C. A. was 15-3 for Grade IX and 17-1 for Grade XI. The mean M. A. was 14-2 for Grade IX and 16-4 for Grade XI according to the California Short Form Test of Mental Maturity. For reasons not yet determined, the girls of this school were at a serious disadvantage on the non-language section of the intelligence test. However, teacher judgment and evaluation of scholastic achievement reveals that the classes were made up of a good number of superior students as well as average and below average students. Therefore the experimental population is considered fairly representative of a typical ninth and eleventh grade population in the regular high school of the average community.

### Analysis of Data

Reliability of test instruments: Chance half reliability coefficients

TABLE I

## RELIABILITY COEFFICIENTS OF CORRELATION FOR VOCABULARY TESTS OF TOTAL POPULATION IN GRADE IX

	Recall		Recognition	
	Uncorrected	Corrected	Uncorrected	Corrected
Nouns	.659	.795	.873	.934
Verbs	.807	.893	.874	.933
Adjectives	.740	.850	.845	.938

TABLE II

## RELIABILITY COEFFICIENTS OF CORRELATION FOR VOCABULARY TESTS OF TOTAL POPULATION IN GRADE XI

	Recall		Recognition	
	Uncorrected	Corrected	Uncorrected	Corrected
Nouns	.607	.7403	.876	.938
Verbs	.585	.738	.815	.898
Adjectives	.685	.813	.873	.932

TABLE III

## COMPARISON OF MEAN TOTAL RAW SCORES — RECALL AND RECOGNITION TEST FORMS FOR TOTAL POPULATION

Test	Form	Grade	No.	Mean	S. E. M	S. D.	Diff.	S. E. Diff.	C. R.
Recall	IX		192	65.31	1.94	26.97			
Recognition	IX		192	151.855	3.21	44.50	86.545	3.75	23.037
Recall	XI		210	91.78	2.0	29.12			
Recognition	XI		210	186.13	3.36	48.65	94.35	3.89	24.25

TABLE IV

## COMPARISON OF MEAN RAW SCORES FOR RECALL AND RECOGNITION FORMS OF NOUN TEST

Test Form	Grade	No. Cases	Mean	S. E. M	S. D.	S. E. Diff.	S. E. Diff.	C. R.
Recall	IX	192	34.208	.973	13.49			
Recognition	IX	192	61.525	1.29	17.90	27.213	1.61	16.45
Recall	XI	210	46.004	1.04	15.11			
Recognition	XI	210	71.190	1.28	18.57	25.196	1.66	15.17

TABLE V

## COMPARISON OF MEAN RAW SCORES FOR RECALL AND RECOGNITION FORMS OF VERB TEST

Test Form	Grade	No. Cases	Mean	S. E. M	S. D.	S. E. Diff.	S. E. Diff.	C. R.
Recall	IX	192	22.456	.79	10.92			
Recognition	IX	192	55.7	1.22	16.92	33.344	1.45	22.92
Recall	XI	210	30.316	.79	11.57			
Recognition	XI	210	68.710	1.32	19.06	38.394	1.53	25.09

TABLE VI

## COMPARISON OF MEAN RAW SCORES FOR RECALL AND RECOGNITION FORMS OF ADJECTIVE TEST

Test		No.	S. E.				S. E.	C. R.
Form	Grade	Cases	Mean	M	S. D.	Diff.	Diff.	
Recall	IX	192	10.842	.414	5.74			
Recognition	IX	192	34.945	1.011	14.02	24.103	1.09	22.09
Recall	XI	210	17.142	.464	6.73			
Recognition	XI	210	47.285	1.17	16.97	30.143	1.58	23.90

TABLE VII

## COEFFICIENT OF CORRELATION BETWEEN THE TOTAL RAW SCORES FOR RECALL AND RECOGNITION FORMS FOR EACH GRADE

Test Form	Grade	No. Cases	Recognition r	Correction for Attenuation
Recall	IX	192	.62 ± .04	.70
Recall	XI	210	.60 ± .05	.72

TABLE VIII

## COEFFICIENT OF CORRELATION BETWEEN THE TOTAL RAW SCORES FOR RECALL AND RECOGNITION FORMS FOR TOP QUARTER OF POPULATION (STUDENTS WITH HIGHEST M. A.)

Test Form	Grade	No. Cases	Recognition r	Correction for Attenuation
Recall	IX	60	.61 ± .07	.69
Recall	XI	53	.52 ± .10	.62

TABLE IX

COEFFICIENT OF CORRELATION BETWEEN THE TOTAL RAW SCORES  
FOR RECALL AND RECOGNITION FORMS FOR LOWER QUARTER OF  
POPULATION (STUDENTS WITH LOWEST M. A.)

Test Form	Grade	No. Cases	Recognition <i>r</i>	Correction for Attenuation
Recall	IX	61	.63 ± .08	.73
Recall	XI	63	.52 ± .11	.39

TABLE X

COEFFICIENTS OF CORRELATION BETWEEN TOTAL RAW SCORE ON  
RECALL FORM AND MENTAL AGE FOR TOTAL POPULATION \*

	Grade IX <i>r</i>	Grade XI <i>r</i>
Total Mental Factors	.477 ± .06	.58 ± .05
Non Language	.298 ± .07	.355 ± .06
Language	.448 ± .06	.634 ± .05

TABLE XI

COEFFICIENTS OF CORRELATION BETWEEN TOTAL RAW SCORE ON  
RECOGNITION FORM AND MENTAL AGE FOR TOTAL POPULATION

	Grade IX <i>r</i>	Grade XI <i>r</i>
Total Mental Factors	.472 ± .06	.518 ± .05
Non Language	.168 ± .07	.241 ± .07
Language	.545 ± .05	.514 ± .05

\*All references to mental age are those obtained from the California Short-Form Test of Mental Maturity (Advanced Form).

based upon 100 cases for each test form showed the coefficients given in Tables I and II.

The high reliability coefficients are to be expected in view of the length of the tests — 115 items for each form of each test. Since only about 12 of the 346 items on recall form had alternative answers, it is not surprising that reliability coefficients for recall form range from .73 to .89 even though the limited achievement on recall form sharply reduces test length below that on recognition form.

Comparison of recall and recognition techniques: — In the vocabulary tests of this study, where identical items have been used in both forms of each test, the mean raw scores on the recognition form of all three tests run from 1.5 to 3 times higher than the mean raw scores on the recall form of the three tests. In order to determine the statistical significance of this difference in mean raw scores or recall and recognition forms, critical ratios have been computed.

Critical ratios for the total raw score and the separate test raw scores are extremely high for both Grade IX and Grade XI. Therefore, the difference in achievement on recall and recognition test forms is statistically significant for the vocabulary tests used in this project.

Examination of the relationship involved in recall and recognition vocabulary abilities is carried still further in the following tables.

The correlation between total raw scores for recall and recognition test forms is positive but only moderately high.

In lieu of the fact that achievement on the total recognition vocabulary is approximately 2.5 times greater than achievement on the total recall vocabulary, a partial explanation of the moderate degree of correlation between the two techniques may be that many low achievers on recall actually are able to make higher scores on recognition forms. In correcting the tests and tabulating scores, it was observed that the students making the very, very low score on the recall form would make a higher score on the recognition form of the same test. On the other hand, a student making a score above the mean on recall seldom made an appreciably lower score on the recognition form of the same test.

Correlation between recall and recognition test forms for students of top and lower quarters are given in Tables VIII and IX.

The low positive correlation of .32 for lower quarter of Grade XI could result from the fact that the advanced students of this grade are slightly more successful on the recognition form of the majority of tests than on the recall form.

Comparison of achievement on separate vocabulary tests. — Since the test items are segregated according to parts of speech and since the different tests are equated in form, style of writing, and length, it is possible to investigate factors relative to parts of speech — nouns, verb, adjectives. The first fact noted in inspecting the scores is the

TABLE XII

## COMPARISON OF MEAN RAW SCORES OF NOUNS AND VERBS — RECALL FORM

Test	Grade	No. Cases	S. E.			Diff.	S. E. Diff.	C. R.
			Mean	M	S. D.			
Nouns	IX	192	34.208	.97	13.49	11.752	1.57	9.39
Verbs	IX	192	22.465	.79	10.92			
Nouns	XI	210	46.004	1.04	15.11	15.688	1.72	11.86
Verbs	XI	210	30.316	.80	11.57			

TABLE XIII

## COMPARISONS OF MEAN RAW SCORES FOR NOUNS AND ADJECTIVES — RECALL FORM

Test	Grade	No. Cases	S. E.			Diff.	S. E. Diff.	C. R.
			Mean	M	S. D.			
Nouns	IX	192	34.208	.97	13.49			
Adjectives	IX	192	10.842	.41	5.74	23.37	1.06	22.0
Nouns	XI	210	46.004	1.04	15.11			
Adjectives	XI	210	17.142	.46	6.73	28.86	1.14	25.27

TABLE XIV

COMPARISONS OF MEAN RAW SCORES FOR NOUNS AND ADJECTIVES  
— RECALL FORM

Test	Grade	No. Cases	Mean	S. E. M	S. D.	Diff.	S. E. Diff.	C. R.
Verbs	IX	192	22.456	.79	10.92			
Adjectives	IX	192	10.842	4.14	5.74	11.614	.94	12.31
Verbs	XI	210	30.316	.80	11.57			
Adjectives	XI	210	17.142	.46	6.73	13.174	.92	14.29

TABLE XV

## COMPARISON OF MEAN RAW SCORES FOR NOUNS AND VERBS — RECOGNITION FORM

Test	Grade	No. Cases	Mean	S. E. M	S. D.	Diff.	S. E. Diff.	C. R.
Nouns	IX	192	61.525	1.29	17.90	6.825	1.78	3.83
Adjectives	IX	192	55.7	1.22	16.92			
Nouns	XI	210	71.190	1.28	18.57	2.48	1.83	1.35
Adjectives	XI	210	68.71	1.315	19.06			

TABLE XVI

COMPARISON OF MEAN RAW SCORES FOR NOUNS AND ADJECTIVES-  
RECOGNITION FORM

Test	Grade	No. Cases	S. E.			S. E. Diff.	C. R.
			Mean	M	S. D.		
Nouns	IX	192	61.525	1.29	17.90		
Adjectives	IX	192	34.945	1.01	14.02	26.580	1.64
Nouns	XI	210	71.190	1.28	18.57		
Adjectives	XI	210	47.285	1.17	16.97	23.905	1.73

TABLE XVII

COMPARISON OF MEAN RAW SCORES FOR NOUNS AND ADJECTIVES-  
RECOGNITION FORM

Test	Grade	No. Cases	S. E.			S. E. Diff.	C. R.
			Mean	M	S. D.		
Verbs	IX	192	55.7	1.22	16.92		
Adjectives	IX	192	34.945	1.22	14.02	20.755	1.73
Verbs	XI	210	68.710	1.31	19.06		
Adjectives	XI	210	47.285	1.17	16.97	21.425	1.76

TABLE XVIII

COEFFICIENT OF CORRELATION BETWEEN RAW SCORES ON RECALL FORM OF THE SEPARATE TESTS FOR GRADE IX

Test	Verb Recall r	Correction for Attenuation	Adjective Recall r	Correction for Attenuation
Noun Recall	.757 $\pm$ .03	.87	.583 $\pm$ .05	.70
Verb Recall			.606 $\pm$ .06	.70

TABLE XIX

COEFFICIENTS OF CORRELATION BETWEEN RAW SCORES ON RECALL FORM OF THE SEPARATE TESTS FOR GRADE XI

Test	Verb Recall r	Correction for Attenuation	Adjective Recall r	Correction for Attenuation
Noun Recall	.652 $\pm$ .04	.80	.5598 $\pm$ .05	.72
Verb Recall			.538 $\pm$ .05	.70

TABLE XX

COEFFICIENT OF CORRELATION BETWEEN RAW SCORES ON RECOGNITION FORM OF THE SEPARATE TESTS FOR GRADE IX

Test	Verb Recognition r	Correction for Attenuation	Adjective Recognition r	Correction for Attenuation
Noun Recognition	.715 ± .04	.78	.794 ± .03	.84
Verb Recognition			.741 ± .03	.80

TABLE XXI

COEFFICIENT OF CORRELATION BETWEEN RAW SCORES ON RECOGNITION FORM OF THE SEPARATE TESTS FOR GRADE XI

Test	Verb Recognition r	Correction for Attenuation	Adjective Recognition r	Correction for Attenuation
Noun Recognition	.716 ± .03	.79	.657 ± .04	.71
Verb Recognition			.638 ± .04	.70

difference in achievement for the various tests. This difference is evaluated in terms of critical ratios also.

Sex differences in vocabulary achievement: No attempt was made to equate boys and girls according to mental age in this study and investigation of sex difference in vocabulary ability is carried no further than to report that the boys of the experimental population maintain a higher mean raw score than girls.

### Conclusions:

#### I. Comparison of recall and recognition test forms

##### A. Achievement in raw scores

1. Total vocabulary score (combined nouns, verbs, adjectives)
  - a. The average high school student of Grades IX and XI was able to recall and write only 45 per cent of the words he was able to recognize and associate correct meanings with.
  - b. Bright students of experimental population were able to recall and write, on the average, approximately 52 per cent of the words which they were able to recognize.
  - c. Dull students were able to recall and write, on the average approximately 40 per cent of the words they were able to recognize.

2. Separate tests

The average high school student of Grade IX and XI was able to recall and write approximately

60 per cent of the noun words which he was able to recognize  
42 per cent of the verb words which he was able to recognize  
32 per cent of the adjective words which he was able to recognize.

##### B. Correlations

1. The coefficients of correlation between achievement on recall and recognition test forms was positive but only moderately high. Therefore, it is considered that
  - a. The two vocabulary test forms have only a limited number of factors in common.
  - b. Achievement on either test form cannot be used to predict achievement on the other with any degree of certainty.
  - c. The correlation between achievement on recall and recognition test forms is no higher for the bright students than it is for the total population.
2. The correlation between achievement or recognition test form and mental age was only moderately high, and the recognition test form correlated as well with mental age as the recall form did.

**II. Comparison of achievement on separate tests.****A. Achievement in raw scores**

1. The highest raw score occurred for nouns and the lowest for adjectives.
2. The mean raw score for nouns was 3 times greater than the mean raw score for adjectives for both recall and recognition test forms.
3. The mean raw score for verbs was 2 times greater than the mean raw score for adjectives on both recall and recognition test forms.
4. The mean raw score for nouns was 1.5 times greater than for verbs on the recall test form, but, on the recognition test form, the difference in raw scores for nouns and verbs are negligible.

**III. Sex differences in vocabulary scores**

There were slight differences in mean raw scores for girls and boys. Those differences were consistently in favor of the boys.

**Summary**

1. Vocabulary scores vary according to test form or mental and language activity, part of speech or syntactic function of words, and grade level in school.
2. Recall-completion and recognition-multiple choice techniques for testing word meaning appear to have only a limited number of factors in common. Therefore, it is possible both recall and recognition activities are necessary if a vocabulary test is to be an adequate evaluation of vocabulary ability.
3. Coefficients of correlation for recognition - multiple choice test form and mental age are approximately the same as the coefficients of correlation for recall-completion test form and mental age.
4. At the high school level, boys appear to hold their own with girls in vocabulary knowledge.

## BIBLIOGRAPHY

1. Brownell, William A. "Some Neglected Criteria for Evaluating Classroom Tests", Appraising The Elementary-School Program, Sixteenth Yearbook, The National Elementary School Principal, Bulletin of the Department of Elementary School Principals (Washington, D. C. The National Education Association, 1937), pp. 485-92.
2. Courtney, Douglas; Buckman, Margaret E.; and Durrell, Donald. "Multiple Choice Recall Versus Oral and Written Recall", Journal of Educational Research XXXIX (February 1946), pp. 458-61.
3. Cronbeck, Lee J. "Analysis of Techniques for Diagnostic Vocabulary Testing", Journal of Educational Research XXXVI (November 1942), pp. 206-17.
4. Crosscup, Richard B. A Survey and Analysis of Methods and Techniques for Fostering Growth of Meaning Vocabulary, Unpublished Master's Thesis, Boston University, School of Education, 1940.
5. Durrell, Donald D. "Language and the Higher Mental Processes" Review of Educational Research XIII (April 1943), pp. 110-14.
6. Gray, William S. and Holmes, Eleanor. The Development of Meaning Vocabularies in Reading, An Experimental Study. Publications of the Laboratory School of the University of Chicago, No. 6 (Chicago: University of Chicago Press, 1938).
7. Kelley, Victor H. An Experimental Study of Certain Techniques for Testing Word Meaning, University of Iowa Studies in Education IX No. 1 (Iowa City: University of Iowa, 1934), pp. 50.
8. Morris, Charles. Signs, Language and Behavior (New York: Prentice-Hall, 1946), pp. 365.
9. Seegar, J. Conrad. "Recent Research in Vocabulary Development" Elementary English Review, XXIII (February 1946), pp. 61-8.
10. Smith, Mary Katherine. "Measurement of the Size of General English Vocabulary Through the Elementary Grades and High School" Genetic Psychology Monograph XXIV (1941), pp. 313-45.
11. Varney, Elinor C. An Experiment to Evaluate Some Techniques for

Measuring the Knowledge of Word Meanings in the Fourth, Fifth, and Sixth Grades. Unpublished Master's Thesis, Boston University, School of Education, 1945.

12. Watts, A. F. The Language and Mental Development of Children (Boston: D. C. Heath and Co., 1944), pp. 354.
13. Wilking, Steven Vincent. The Construction and Evaluation of a Measure of Reading Vocabulary, Unpublished Master's Thesis, Boston University School of Education, 1940.
14. Zipf, George Kingsley. The Psycho-Biology of Language. (Boston: Houghton-Mifflin Co., 1935), pp. 336.

# PREDICTION OF SUCCESS IN COLLEGE MATHEMATICS AT WASHBURN UNIVERSITY

WILLIAM F. SEIGLE

Washburn Municipal University  
Topeka, Kansas

STUDENTS HAVE often been placed in courses or they have chosen them without any knowledge of their fitness for these courses. Some of the students have chosen wisely, while others have failed or have withdrawn from the courses. The results at Washburn University, Topeka, Kansas, have not been entirely satisfactory, because a too large percent of those enrolling in mathematics courses either have failed or have withdrawn from the courses.

It seems reasonable to suppose that some of these failures could have been avoided and that many students could have benefited more from courses that were better suited to their abilities. It was felt that guidance was necessary to direct the capable student into choosing mathematics and guidance was also necessary to direct the student who was almost sure to fail in mathematics into some more suitable field. If an inexpensive and time-saving device would successfully predict the achievement of these students, much could be done in aiding them to choose the right course.

This study was conducted at Washburn Municipal University, Topeka, Kansas. It covers the years of 1945, 1946, 1947, 1948, and 1949. The study includes the complete records of all 1205 students who entered the mathematics department.

At Washburn University every student is required to take a minimum of five semester hours in at least six of the seven fields of work that are offered by the University. As a result, the beginning courses have a great many students who never take more than the maximum five hours in the field of mathematics. The students who take advanced work in this department are much more homogeneous in their background and interests than those students in the beginning courses.

The general intelligence test has been considered valuable in determining a student's ability to do work of college caliber, and within recent years emphasis has been placed on the development and use of special aptitude tests for the different subjects rather than a general academic ability test. The problem of placement becomes one of finding those who are unfitted for the work in a special field and guiding them into a more suitable field. For predictive purposes the first consideration might be the previous records of the students.

The studies previously made, however, seem to point to the fact

that grades in college are dependent, at least in part, on factors other than intelligence. In attempts to measure these factors, recourse has been made to personality tests of various sorts. While it is generally conceded that personality factors do enter into the picture, the attempts to measure these traits for predictive purposes have met with no striking success.

Subjective elements and different standards of work make high school records unreliable as a basis for predicting success in college mathematics. There seems to be a fairly definite correlation between scores on English tests and grades in college mathematics. The same thing can also be said for scores on reading comprehension. General intelligence tests have a definite correlation with success in mathematics, although the correlation is not high enough to allow it to be used for individual prognosis. Special mathematical aptitude tests give a better basis for prognosis than do general intelligence tests. Personality tests seem to offer possibilities for development as a supplement to the intelligence and special aptitude tests. However, in their present state of development, they have been of little value.

The author made this investigation to find some of the factors that are predictive of success in college mathematics at Washburn. He also wanted to find the relative prognostic value of the Washburn Entrance Mathematics Test, high school grade average, the number of units of high school mathematics, the American Council of Education Psychological Test, the American Council of Education English Test, and combinations of these instruments.

Scores on each of these factors were obtained for all students enrolled in the mathematics department. It was felt that the grades as given by the teachers in this department would probably be the best available criterion of success. Since nearly all of the teachers in this department give the same tests, and since all of them give the same final examination, the author felt that the grades would be fairly consistent throughout the department.

The Washburn Entrance Mathematics Test is composed of forty items and thirty minutes is ample time for its administration. The validity of it as a prognostic test was determined by the coefficient of correlation between scores on it and grades in college algebra. This coefficient was .75. The reliability of the test was .953 as determined by the application of the split-half technique and the Spearman-Brown prophesy formula.

#### Prediction of Success in College Algebra

Beginning students are permitted to take Mathematics I which is intermediate algebra and college algebra, or they can take Mathematics 9 which is intermediate algebra, college algebra, and trigonometry.

Both are five hour courses. Several of the students that took Mathematics I continued with their mathematics by taking a course in trigonometry.

In order to enable the counselors in Mathematics to do a better job of guiding students, it was hoped that an inexpensive and time-saving device could be developed for the prediction of success in these courses.

The different prognostic factors in this part of the study were:

1. Scores on the Washburn Entrance Mathematics Test
2. Units of high school mathematics
3. High school grade point average
4. Q-scores on the A. C. E. Psychological Test
5. L-scores on the A. C. E. Psychological Test
6. Total scores on the A. C. E. Psychological Test
7. Reading comprehension scores on the A. C. E. English Test
8. Mechanics of expression scores on the A. C. E. English Test
9. Effectiveness of expression scores on the A. C. E. English Test
10. Total scores on the A. C. E. English Test

The coefficients of correlation between the scores of each of these factors and the grades in college algebra were computed by means of the Pearson Product-Moment method. Also, the coefficients of correlation between scores on each of the prognostic factors and each of the other prognostic factors were computed. The best multiple correlation was determined by means of the Wherry-Doolittle Method. Other coefficients of multiple correlation were found by the partial method. These are found in Table I.

In the part of the study that dealt with the prediction of success in college algebra the multiple correlation was found to be .82. From this it was found that the battery of tests selected accounted for 66.8 percent of the variance of the criterion. The Washburn Entrance Mathematics Test accounted for 48.4 percent, the high school grade average accounted for 12.5 percent, the Q-score on the A. C. E. Test accounted for 1.9 percent, and the total score on the A. C. E. Test accounted for .5 percent of the variance.

#### Prediction of Success in Mathematics Through Trigonometry.

Students who take Mathematics I then take a course in trigonometry and complete the same material as those who enroll in Mathematics 9. The main difference in these two programs is the number of semester hours required. A student taking Mathematics 9 completes algebra and trigonometry by taking five semester hours, while Mathe-

matics I followed by a course in trigonometry requires nine semester hours.

Ten different prognostic factors were set up. These were:

1. Scores on the Washburn Entrance Mathematics Test
2. Semesters of high school mathematics.
3. Grade point average in all high school subjects.
4. Q-score on the American Council of Education Psychological Test.
5. L-score on the American Council of Education Psychological Test.
6. Total Score on the American Council of Education Psychological Test.
7. Total Reading Comprehension score on the American Council of Education English Test.
8. Mechanics of Expression score on the American Council of Education English Test.
9. Effectiveness of Expression score on the American Council of Education English Test.
10. Total English score on the American Council of Education English Test.

The coefficients of correlation between the scores of each of these factors and the grades in college mathematics through trigonometry were computed by means of the Pearson Product-Moment method. Also, the coefficients of correlation between the scores on each of the prognostic factors and each of the other prognostic factors were computed. The best coefficients of multiple correlation were computed by using the partial method. Regression equations for each were also computed. These are found in Table II.

In this part of the study dealing with the prediction of success in college mathematics through trigonometry the multiple correlation coefficient was found to be .7314. The battery of tests selected accounted for 53.6 percent of the variance of the criterion. The Washburn Entrance Test accounted for 25.3 percent of the variance and the high school grade average accounted for a similar amount. The English Test accounted for about 3 percent of the variance.

Prediction of Success in Analytic Geometry (before any college mathematics courses have been taken.)

In this part of the study the author presents the problem dealing with the prediction of success in analytic geometry before the student has taken any courses in college mathematics.

The American Council of Education Psychological Test, the American Council of Education English Test, and the Washburn Entrance Mathematics Test has been given to all students who enrolled in the Mathematics Department in the years 1945, 1946, 1947, 1948, 1949.

TABLE I

## SUMMARY OF MULTIPLE CORRELATION COEFFICIENTS FOR COLLEGE ALGEBRA

Predictive Variables	Correlation Coefficient	Standard Error of Estimate	Regression Equation
R <sub>1.42537</sub>	.8171	1.58	X <sub>1</sub> = .2405 X <sub>2</sub> + .1370 X <sub>3</sub> + .0232 X <sub>5</sub> + .0016 X <sub>7</sub> - 1.4697
R <sub>1.42</sub>	.801	1.64	X <sub>1</sub> = .304 X <sub>4</sub> + 1.259 X <sub>2</sub> - 2.135
R <sub>1.43</sub>	.757	1.79	X <sub>1</sub> = .278 X <sub>4</sub> + .183 X <sub>3</sub> - .437
R <sub>1.45</sub>	.763	1.77	X <sub>1</sub> = .265 X <sub>4</sub> + .042 X <sub>5</sub> - 1.144
R <sub>1.47</sub>	.759	1.78	X <sub>1</sub> = .264 X <sub>4</sub> + .017 X <sub>7</sub> - 1.1845

TABLE II

## SUMMARY OF MULTIPLE CORRELATION COEFFICIENTS FOR COLLEGE MATHEMATICS THROUGH TRIGONOMETRY \*

Predictive Variables	Correlation Coefficient	Standard Error of Estimate	Regression Equation
R <sub>1.42 11</sub>	.7314		
R <sub>1.42</sub>	.693	1.63	X <sub>1</sub> = .139 X <sub>4</sub> + 1.28 X <sub>2</sub> + .4575
R <sub>1.2 11</sub>	.645	1.74	X <sub>1</sub> = .173 X <sub>4</sub> + .021 X <sub>7</sub> - .4363
R <sub>1.4 11</sub>	.650	1.73	X <sub>1</sub> = .168 X <sub>4</sub> + .020 X <sub>11</sub> - 1.197

From these scores eleven different prognostic factors were set up. These were:

- Variable 1. Grades in Analytic Geometry
- Variable 2. High School Grade Average
- Variable 3. Semesters of High School Mathematics
- Variable 4. Scores on Washburn Entrance Mathematics Tests
- Variable 5. "Q" Scores on the A. C. E. Psychological Test
- Variable 6. "L" Scores on the A. C. E. Psychological Test
- Variable 7. Total Scores on the A. C. E. Psychological Test
- Variable 8. Reading Comprehension Score
- Variable 9. Mechanics of Expression Score
- Variable 10. Effectiveness of Expression Score
- Variable 11. Total English Score.

The coefficients of correlation between the scores of each of these predictive factors and the grades in Analytic Geometry were computed by means of the Pearson Product-Moment method. Also, the coefficients of correlation between the scores on each of the predictive factors and each of the other predictive factors were computed. The best coefficient of multiple correlation was found by the Wherry-Doolittle method. Other coefficients of multiple correlation were found by using the partial method. In order to ascertain whether a combination of fewer predictive variables would be of value, the author computed some other coefficients of multiple correlation. Regression equations for each were also computed. These are given in Table III.

In the part of the study that dealt with the prediction of success in analytic geometry at the time that the student entered college the multiple correlation was found to be .668. It was found that the battery of tests selected accounted for 44.6 percent of the variance of the criterion. The Washburn Entrance Mathematics Test accounted for 18.6 percent of the variance, the high school grade average accounted for 21.7 per cent of the variance and the L-score on the A. C. E. Psychological Test accounted for 4.3 percent.

#### Prediction of Success in Analytic Geometry (after some college mathematics courses have been taken.)

Now, we come to the second part of our prediction problem, namely predicting success in analytic geometry after they have had some college mathematics. Often students have taken college mathematics through trigonometry and they are faced with the problem of whether they should continue in the field of mathematics. This part of the study is devoted to that problem.

The American Council of Education Psychological Test, the American Council of Education English Test, and the Washburn Entrance Mathematics Test had been given to all students who enrolled

in the Mathematics Department in the years of 1945, 1946, 1947, 1948, 1949. Their high school grade average was known, the numbers of semesters of high school mathematics, their grades in college algebra and trigonometry, and the grades of that group who took analytic geometry were available.

From these scores that were obtained, twelve different prognostic factors were set up. These were:

- Variable 1. Grade in Analytic Geometry
- Variable 2. Math Grade through Trigonometry
- Variable 3. High School Grade Average
- Variable 4. Semesters of High School Mathematics
- Variable 5. Score on Washburn Entrance Mathematics Test
- Variable 6. "Q" Score on A. C. E. Psychological Test
- Variable 7. "L" Score on A. C. E. Psychological Test
- Variable 8. Total Score on A. C. E. Psychological Test
- Variable 9. Reading Comprehension Score
- Variable 10. Mechanics of Expression Score
- Variable 11. Effectiveness of Expression Score
- Variable 12. Total English Score

The coefficients of correlation between the scores of each of these predictive factors and the grades in analytic geometry were computed by means of the Pearson Product-Moment method. Also, the coefficients of correlation between the scores on each of the predictive factors and each of the other predictive factors were computed. The best coefficient of multiple correlation was computed by using the Wherry-Doolittle method.

In order to ascertain whether a combination of fewer predictive variables would be of value, the author computed some other coefficients of multiple correlation. Regression equations for each were also computed. These are found in Table IV.

In this part of the study that dealt with the prediction of success in analytic geometry after trigonometry had been taken, the coefficient of multiple correlation was found to be .7858. It was determined that the battery of tests selected accounted for 61.8 percent of the variance of the criterion. The grades made in college mathematics through trigonometry accounted for 47.7 percent of the variance, 9.9 percent was accounted for by the high school grade average, and the other 4.2 percent was accounted for by the Washburn Entrance Mathematics Test.

#### Prediction of Success in Calculus (before any college mathematics courses have been taken)

In this part of the study the author presents the problem dealing with the prediction of success in calculus before the student has taken

TABLE III

## SUMMARY OF MULTIPLE CORRELATION COEFFICIENTS FOR COLLEGE MATHEMATICS THROUGH ANALYTICAL GEOMETRY

Predictive Variables	Correlation Coefficient	Standard Error of Estimate	Regression Equation
R <sub>1.357</sub>	.668	1.73	X <sub>1</sub> = 1.50 X <sub>3</sub> + .123 X <sub>5</sub> + .014 X <sub>7</sub> + .078
R <sub>1.35</sub>	.663	1.75	X <sub>1</sub> = 1.68 X <sub>3</sub> + .132 X <sub>5</sub> - 0.35
R <sub>1.37</sub>	.577	1.91	X <sub>1</sub> = 1.79 X <sub>3</sub> + .037 X <sub>7</sub> + 0.96
R <sub>1.37</sub>	.560	1.94	X <sub>1</sub> = .146 X <sub>5</sub> + .042 X <sub>7</sub> - 0.18

TABLE IV

## SUMMARY OF MULTIPLE CORRELATION COEFFICIENTS FOR COLLEGE MATHEMATICS THROUGH ANALYTICAL GEOMETRY

Predictive Variables	Correlation Coefficient	Standard Error of Estimate	Regression Equation
R <sub>1.235</sub>	.7858	1.41	X <sub>1</sub> = .680 X <sub>2</sub> + .690 X <sub>3</sub> + .028 X <sub>5</sub> - .412
R <sub>1.23</sub>	.782	1.46	X <sub>1</sub> = .745 X <sub>2</sub> + .656 X <sub>3</sub> - .01
R <sub>1.25</sub>	.771	1.49	X <sub>1</sub> = .737 X <sub>2</sub> + .028 X <sub>5</sub> + .50
R <sub>1.35</sub>	.663	1.75	X <sub>1</sub> = 1.68 X <sub>3</sub> + .132 X <sub>5</sub> - .35

any courses in college mathematics.

The American Council of Education Psychological Test, the American Council of Education English Test, and The Washburn Entrance Test had been given to all students who enrolled in the mathematics department in the years 1945, 1946, 1947, 1948, and 1949.

From the scores that were obtained, eleven different prognostic factors were set up. These were:

- Variable 1. Grades in Calculus
- Variable 2. High School Grade Average
- Variable 3. Semesters of High School Mathematics
- Variable 4. Scores on Washburn Entrance Mathematics Test
- Variable 5. "Q" Scores on A. C. E. Psychological Test
- Variable 6. "L" Scores on A. C. E. Psychological Test
- Variable 7. Total Scores on A. C. E. Psychological Test
- Variable 8. Reading Comprehension Score
- Variable 9. Mechanics of Expression Score
- Variable 10. Effectiveness of Expression Score
- Variable 11. Total English Score

The coefficient of correlation was computed between the scores on each of these by the Pearson Product-Moment Method. Also, the coefficients of correlation were computed between the scores on each of the prognostic factors and each of the other prognostic factors. The best coefficient of multiple correlation was determined by the Wherry-Doolittle Method. In order to ascertain whether a combination of fewer predictive variables would be of any value, the author computed some other coefficients of multiple correlation. Regression equations for each were also computed. A summary of these is given in Table V.

In the part of the study that dealt with the prediction of success in calculus before any college mathematics had been taken, the coefficient of multiple correlation was found to be .626. It was determined that the battery of tests selected accounted for 39 percent of the variance of the criterion. The Washburn Entrance Mathematics Test accounted for 17 percent of the variance, the high school grade average accounted for 10 percent of the variance, while the English test accounted for 12 percent of the variance.

#### Prediction of Success in Calculus (after some college mathematics courses have been taken)

Now, we come to the predicting success in calculus after students have taken some college mathematics. Often students have taken college mathematics through analytic geometry and are faced with the problem of whether they should continue in the field of mathematics or not. This part of the study is devoted to that problem.

The American Council of Education Psychological Test, the

TABLE V

## SUMMARY OF MULTIPLE CORRELATION COEFFICIENTS FOR SUCCESS IN CALCULUS BEFORE TAKING ANY COLLEGE MATHEMATICS

Predictive Variables	V Correlation Coefficient	Standard Error of Estimate	Regression Equation
R <sub>1.6134</sub>	.626	2.01	X <sub>1</sub> = 1.40 X <sub>6</sub> + .027 X <sub>13</sub> + .812X <sub>4</sub> - 7.84
R <sub>1.613</sub>	.602	2.09	X <sub>1</sub> = .158 X <sub>6</sub> + .038 X <sub>13</sub> + 4.49
R <sub>1.46</sub>	.592	2.12	X <sub>1</sub> = 1.49 X <sub>4</sub> + .150 X <sub>6</sub> - 1.33
R <sub>1.413</sub>	.561	2.18	X <sub>1</sub> = 1.17 X <sub>4</sub> + 0.36 X <sub>3</sub> - 1.63

TABLE VI

## A SUMMARY OF MULTIPLE CORRELATION COEFFICIENTS FOR SUCCESS IN CALCULUS AFTER TAKING ANALYTIC GEOMETRY

Predictive Variables	Correlation Coefficient	Standard Error of Estimate	Regression Equation
R <sub>1.263</sub>	.81	1.46	X <sub>1</sub> = .835X <sub>2</sub> + .074X <sub>6</sub> + .101 X <sub>3</sub> - 2.59
R <sub>1.23</sub>	.824	1.49	X <sub>1</sub> = .780X <sub>2</sub> + .268 X <sub>3</sub> - 1.3197
R <sub>1.26</sub>	.828	1.47	X <sub>1</sub> = 091X <sub>2</sub> + .084 X <sub>6</sub> - 2.7639
R <sub>1.36</sub>	.718	1.83	X <sub>1</sub> = .732 X <sub>3</sub> + .084 X <sub>6</sub> - 1.8177

American Council of Education English Test, and the Washburn Entrance Mathematics Test had been given to all students who enrolled in the mathematics department in the years of 1945, 1946, 1947, 1948, 1949. Their high school grade average was known, the number of semesters of high school mathematics, their grades in college algebra and trigonometry, and in analytic geometry, and the grades of that group which took calculus were available.

From these scores that were obtained, thirteen different prognostic factors were set up. These were:

- Variable 1. Grade in Calculus
- Variable 2. Grade in Analytic Geometry
- Variable 3. Grade through Trigonometry
- Variable 4. High School Grade Average
- Variable 5. Semesters of High School Mathematics
- Variable 6. Scores on Washburn Entrance Mathematics Test
- Variable 7. "Q" Score on A. C. E. Psychological Test
- Variable 8. "L" Score on A. C. E. Psychological Test
- Variable 9. Total Score on A. C. E. Psychological Test
- Variable 10. Reading Comprehension Score
- Variable 11. Mechanics of Expression Score
- Variable 12. Effectiveness of Expression Score
- Variable 13. Total English Score

The coefficient of correlation was computed between the scores of each of these factors and the grades in calculus by means of the Pearson Product-Moment method. Also, the coefficients of correlation were computed between the scores on each of the prognostic factors and each of the other prognostic factors. The best coefficient of multiple correlation was found by means of the Wherry-Doolittle Method. In order to ascertain whether a combination of fewer predictive variables would be of value, the author computed some other coefficients of multiple correlation. Regression equations for each were also computed. A summary of these is given in Table VI.

In the part of the study that dealt with the prediction of success in calculus after analytic geometry had been taken, the coefficient of multiple correlation was found to be .831. From this it was found that the battery of variables selected accounted for 69 percent of the variance of the criterion. Grades in analytic geometry accounted for 54 percent of the variance, grades in college mathematics through trigonometry accounted for 9 percent of it, while the score on the Washburn Entrance Mathematics Test accounted for 6 percent of it.

### Conclusions

- (1) The Washburn Entrance Mathematics Test proved to have value for the prognosis of grades in college mathematics at Washburn

University for the groups studied.

(2) The reliability of the Washburn Entrance Mathematics Test is quite satisfactory even though it is a short test and requires only thirty minutes to administer.

(3) The Washburn Entrance Mathematics Test is the best single predictor of success in college algebra that was found in this study. It was not as effective in predicting success in trigonometry, analytic geometry, and calculus as it was in predicting success in college algebra, yet in all of these areas it contributed something to the coefficient of multiple correlation.

(4) The high school grade average was the second best predictor of success in college mathematics if the prediction was made before any college mathematics had been taken. In the case of the prediction of success in analytic geometry at the time of the student's entrance to college, it was a better predictive factor than the Washburn Entrance Mathematics Test.

(5) The best predictive factor after some college mathematics had been taken was the grade made in the previous mathematics course. In the case of analytic geometry, the grade made in college mathematics through trigonometry was the best predictive factor. The grade made in analytic geometry was the best predictor of success in calculus.

(6) The amount of high school mathematics taken by the student seemed to play a very small part in his success in other mathematics courses after his first course in college mathematics.

(7) The coefficient of correlation between the battery selected and grades in analytic geometry was .6623, while for calculus it was .6183. These are large enough to be of definite value in group prediction and fairly good for individual prediction.

(8) The coefficient of correlation between the battery selected and grades in college algebra was .8171. The coefficient of correlation in the case of analytic geometry after trigonometry had been taken was .7818, while in the case of calculus after analytic geometry had been taken was .8212. All three of these coefficients are large enough to be of very definite value in group prediction and are exceedingly good for individual prediction.

# A DESCRIPTIVE ANALYSIS OF 248 NON-HIGH-SCHOOL GRADUATES ADMITTED TO THE UNIVERSITY OF UTAH

DEAN C. ANDREW  
Southern State College  
Magnolia, Arkansas

## Purpose

DURING World War II, and at the conclusion of hostilities, numerous colleges received applications from veterans who had not completed high school. Many had had their high school education interrupted for various reasons beyond their control, but were capable of doing college work regardless of credit deficiencies. The purpose of this study was to analyze some of the characteristics of such a group of students admitted to the University of Utah. The assumption throughout the study was that the more we know about a student the better able we are to serve him while in college. Also it was felt that the practice of admitting non-high-school graduates was not a temporary thing confined to the World War II period, but will be a continuous thing as long as world tension exists.

## Description of Sample

The student data used in this study were obtained from the registrar's files at the University of Utah. The experimental group used in this investigation was limited to (1) non-high-school graduates who had been admitted to the University of Utah between the fall quarter 1945 and summer quarter 1950; (2) male students; and (3) those who had completed at least one quarter of college work. A sample of 248 non-high-school graduates was obtained for the study. Table I contains information concerning the number of non-high-school graduates enrolled per quarter.

## Characteristics of Non-High-School Graduates

One might want to know many things concerning the background, age, health, training, etc., of the atypical student. The admissions officer certainly desires to ascertain the extent of the non-graduates' previous training in order to evaluate entering students in terms of the policies and procedures of the institution. The description in this study will be in terms of personal characteristics, academic preparation, and psychometric test results. In addition the achievements of the non-high-school graduates while in college will be pointed out.

Table II contains descriptive data regarding some of the personal

TABLE I

## NON-HIGH-SCHOOL GRADUATES ENROLLED BY QUARTERS

Year	Fall	Winter	Spring	Summer	Total
1945-46	18	53	63	25	159
1946-47	41	65	49	21	176
1947-48	62	64	54	29	209
1948-49	61	72	65	37	235
1949-50	76	78	70	22	246

Mean number enrolled each quarter = 51.25

TABLE II

## SOME CHARACTERISTICS OF 248 NON-HIGH-SCHOOL GRADUATES ADMITTED TO THE UNIVERSITY OF UTAH

Variable	Mean	S. D.
1. No. of quarters completed in college	4.13	3.56
2. Age at time of college entrance	23.32	5.10
3. Period of time lapsing between last secondary school attendance and admission to the University of Utah	4.06	3.22
4. No. of high school basic units earned *	7.78	3.11
5. No. of high school units earned	11.22	4.12
6. High school grade point averages **	1.05	.56

\*The high school basic subjects referred to are: English, algebra, plane geometry, solid geometry, trigonometry, general science, physiography, botany, physiology, zoology, physics, chemistry, astronomy, United States history, general history, economics, sociology, psychology, French, German, Latin, and Greek.

\*\*The grade-point average is a ratio of honor points divided by quarter hours taken. The honor points are divided on the following basis: "A"=3, "B"=2, "C"=1, "D"=0, "E"=-1. However, in figuring high school grade-point average, a grade of "E" was not considered because it is not accepted for admissions.

characteristics and academic preparation of the non-high-school graduates.

The average number of quarters attended by the non-high-school graduates is 4.13 and varies from one to eighteen quarters. About 27 per cent of the students attended only one quarter, while nearly 61 per cent of the group only completed three quarters or one academic year. There were five people who attended more than twelve quarters, which is ordinarily sufficient for graduation. This was probably due to the fact that they were required to make up their high school deficiencies before they were allowed to matriculate in the University.

Age at time of entrance into the university varied from 16 to 49 with an average of 23.32 years. Approximately two-thirds of the students admitted fall in the age range from twenty to twenty-five. With the mean age at 23.32, it would appear that the non-high-school graduate is approximately five years older than the average freshman.

The average number of years out of high school was 4.06 with some students going directly from high school into college while one had been out as long as thirty-two years. Nearly one-fourth of these students had enrolled in the university before any great length of time had elapsed. This might indicate that a number of them spent less than two years in military service and thus were able to reenter school without a long period of interruption.

One thing that a college admissions officer is greatly concerned about is the number of high school basic units the person has completed. At the University of Utah, ten basic units are required for entrance. The range of basic units presented for entrance by this group was from zero to fifteen with the average being 7.78. The average non-high-school graduate was deficient 2.22 basic units. Nearly seventy-three per cent of the total group lacked the ten required basic units.

Besides presenting ten basic units, an applicant must present an additional five high school elective units for entrance at the University of Utah, making a total of 15 units. The mean number of units presented by the non-high-school graduate was 11.22, thus indicating a deficiency of 3.78 units. Approximately 79 per cent of the total group presented less than a total of fifteen units for entrance. Assuming that a high school student carries approximately 4.5 units per year, the average non-high-school graduate attended high school for 2.49 years.

In addition to knowing the quantity of high school work taken, it is also very important to know the quality of the work completed. The grades earned in high school are usually more important than the number of units completed. The mean grade-point average obtained by this group was 1.05. This means that the typical non-high-school graduate earned approximately a "C" grade while in high school.

Table III gives the location by state of the secondary schools attended by the experimental group.

TABLE III

LOCATION BY STATE OF SECONDARY SCHOOLS ATTENDED BY THE  
NON-HIGH-SCHOOL GRADUATES

State	No.	Per Cent
Utah	193	78.78
Idaho	10	4.08
California	9	3.67
Iowa	3	1.22
Oregon	3	1.22
Arizona	2	.82
Colorado	2	.82
Missouri	2	.82
New Jersey	2	.82
New York	2	.82
Texas	2	.82
Washington	2	.82
Canada	1	.41
Florida	1	.41
Illinois	1	.41
Minnesota	1	.41
Mississippi	1	.41
Montana	1	.41
New Mexico	1	.41
Oklahoma	1	.41
Pennsylvania	1	.41
South Africa	1	.41
South Carolina	1	.41
Virginia	1	.41
Wyoming	1	.41
Total	245	100.00

One hundred and ninety-three, or 78.78 per cent, came from secondary schools in Utah. Of the 193 students who came from Utah secondary schools, 137 or 70.98 per cent were from the Salt Lake City school systems. That is to say, over one-half of the entire experimental group had attended a secondary school in Salt Lake City.

The entrance examinations at the University of Utah are the Cooperative General Achievement Tests. In addition to the entrance examinations, some of the non-high-school graduates also took the General Educational Development Tests, High School level. The experimental group performance on these tests is shown in Tables IV and V.

In both of the tests the mean score for the non-high-school graduates is above the average of the norm group in every area except English. The experimental group found the greatest difficulty in the English test of the examinations. It is of interest to note that in four of the G. E. D. Tests, the average score of the experimental group was at the 70th percentile or higher when compared with the normal high-school graduates.

#### College Achievement of the Non-High-School Graduates

After determining some of the personal and high school characteristics of the non-high-school graduates, the next logical step is to find out about his performance in college. Table VI shows the division of the university in which the non-high-school graduate enrolled and the number who graduated in each division;

Because so many non-high-school graduates did not finish the first two years, the majority were enrolled in lower division where a student fills required groups before making a choice of a major. It appears from Table VI that out of the 248 students, approximately ten per cent will or have graduated.

The academic achievement of the non-high-school graduates can be seen in Tables VII, VIII, IX, and X. They contain information concerning total marks and grade-point averages, and first-quarter marks and grade-point averages.

The mean grade-point average for total college work was .615 while the mean first quarter grade-point average for this group was .715. In both instances the mean grade-point average was below that of the general college population at the University and was also below that required for passing. It was found that 206 out of 248 had been on probation one or more times. Eighty-two out of the total number had been dropped from school at least once, while only twenty achieved honors. In determining how many were successful and how many were failures, a grade-point average of 1.00 or better was used. This is equivalent to a "C" grade. Eighty-six students, or 34.68 per cent had below a grade-point average of 1.00 or above while 162, or 65.32 per cent had below a 1.00 grade-point average for total college work. It was also ascertained that of the number who withdrew from school, 147 withdrew

TABLE IV

COOPERATIVE ACHIEVEMENT MEAN SCORES AND PERCENTILES  
FOR THE NON-HIGH-SCHOOL GRADUATES

Test	Mean	Percentile of mean (U. S.)	S. D.	N
Coop. English	49.30	39	9.25	241
Coop. Social Studies	59.25	56	10.45	238
Coop. Nat. Sciences	60.46	53	6.36	237
Coop. Mathematics	55.75	56	6.21	238
Ave. Coop. Standard Scores	55.72	--	6.87	236

TABLE V

G. E. D. MEANSCORES AND PERCENTILES FOR 97 NON-HIGH-SCHOOL  
GRADUATES

Test	Mean	Percentile of mean (U. S.)	S. D.	N
G. E. D. Test 1	48.52	44	8.28	97
G. E. D. Test 2	57.85	79	8.55	97
G. E. D. Test 3	59.85	84	7.75	97
G. E. D. Test 4	57.00	76	8.20	97
G. E. D. Test 4	55.90	73	11.45	97
Ave. G. E. D. Standard Score	55.52	--	7.32	97

TABLE VI

**COLLEGE OF ENROLLMENT FOR THE NON-HIGH-SCHOOL GRADUATES  
AND NUMBER OF GRADUATES IN EACH**

College*	No. Matriculated	Per Cent	No. Graduated
Business	22	8.87	2
Education	13	5.24	3
Engineering	23	9.27	1
Pre-Medicine	9	3.63	0
University	24	9.68	7
Fine Arts	4	1.61	1
Languages	3	1.21	3
Terminal	20	8.07	0
Lower Division	126	50.81	0
Military Sciences	4	1.61	3
Total	248	100.00	20

\*The college of enrollment was established by combining courses and majors where they were similar.

TABLE VII

**DISTRIBUTION OF TOTAL MARKS EARNED IN COLLEGE BY 248 NON-HIGH-SCHOOL GRADUATES**

Grade	No. of Hours	Per Cent
A	1477.0	10.08
B	3481.3	13.76
C	5810.8	39.66
D	1694.0	11.57
E	1814.0	12.38
Incomplete	373.5	2.55
Total	14,650.6	100.00

Mean number of quarter hours carried per student = 59.07

Mean number of hours carried per quarter per student = 14.29

TABLE VIII

## TOTAL COLLEGE GRADE-POINT AVERAGES FOR 248 NON-HIGH-SCHOOL GRADUATES

G. P. A.	Frequency	Per Cent
2.50	3	1.21
2.25	2	.81
2.00	8	3.23
1.75	9	3.62
1.50	14	5.65
1.25	17	6.85
1.00	33	13.31
.75	30	12.10
.50	27	10.89
.25	26	10.48
.00	17	6.85
-.25	18	7.26
-.50	10	4.03
-.75	12	4.84
-1.00	22	8.87
Total	248	100.00

Mean = .615, S. D. = .85

TABLE IX

## DISTRIBUTION OF FIRST QUARTER MARKS EARNED BY 248 NON-HIGH-SCHOOL GRADUATES

Grade	No. of Hours	Per Cent
A	184.5	5.47
B	650.0	19.26
C	1296.5	38.41
D	521.0	15.44
E	627.0	18.58
Incomplete	96.0	2.84
Total	3375.0	100.00

TABLE X

DISTRIBUTION OF FIRST QUARTER COLLEGE GRADE-POINT AVERAGES OF 248 NON-HIGH-SCHOOL GRADUATES

G. P. A.	Frequency	Per Cent
2.50	7	2.82
2.25	4	1.61
2.00	8	3.23
1.75	10	4.03
1.50	17	6.85
1.25	21	8.47
1.00	46	18.55
.75	20	8.06
.50	22	8.87
.25	20	8.07
.00	14	5.65
-.25	12	4.84
-.50	7	2.82
-.75	14	5.65
-1.00	26	10.48
Total	248	100.00

Mean = .715, S. D. = .92

failing while forty-three withdrew passing. A general survey of the above achievement records of this atypical group indicates that more than one-half withdrew failing and were put on probation while in attendance at the University.

### Summary

The typical male non-high-school graduate admitted to the University of Utah during 1945-50 was about 23 years of age and had completed approximately two and one-half years of high school. While in high school he had earned nearly eight basic and eleven total units of credit with approximately a "C" grade. The typical entrant had been out of high school for about four years prior to his entrance into college, and on standardized achievement tests he scored above average in all areas except English. While in college he earned an average grade of -C and was placed on probation at least once. He would probably remain in college no longer than two academic years.

# A PRELIMINARY STUDY OF THE GROUP ATTITUDES OF JUNIOR AND SENIOR HIGH SCHOOL STUDENTS TOWARDS MATHEMATICS

WILLIAM H. MALONE

Lowell, Massachusetts

EUGENE L. FREEL

North Adams, Massachusetts\*

WHILE SOME authors declare that one's verbal attitudes may not actually measure what one will do when he encounters a certain object or situation, and, as such, attitude questionnaires merely measure verbal attitudes toward symbolic objects or situations,<sup>1</sup> it would seem plausible that they do represent the current, conscious attitude of the person, in the absence of some concrete explanation why they do not. It is, of course, possible that they do not represent an unconscious attitude, and other techniques would appear necessary to explore that area.

A great deal has been written about attitudes, and there are many definitions of the term. Perhaps the one most accepted is that of Gordon Allport which reads:

An attitude is a mental and neural state of readiness, organized through experience, exerting a directive or dynamic influence upon the individual's response to all objects and situations with which it is related.<sup>2</sup>

Sherif and Cantril say that "many social attitudes are formed through verbal judgments of adults", and as such are "affectionately charged", that is, "formed in relation to social values ~~which in themselves~~ which in them-selves are standardized affective fixations."<sup>3</sup>

\* State Teachers College at Lowell, Mass., and at North Adams, Mass.

1. La Piere and Farnsworth. Social Psychology 3rd ed., (New York: McGraw Hill, 1949), p. 117.
2. Gordon Allport. "Attitudes," in Carl Murchison (ed.) Handbook of Social Psychology (Worcester, Mass.: Clark University Press, 1935), pp. 798-844.
3. Steuart Henderson Britt. Selected Readings in Social Psychology (New York: Rinehart and Co., 1950), p. 135.

A recent statement reads, "A wholesome attitude toward arithmetic is promoted when the child feels (1) that arithmetic is useful, (2) that success for him is possible, and (3) that the teacher is enthusiastic about the subject."<sup>4</sup>

In order to focus our inquiry on the above, we have formulated these broad questions:

1. Are the attitudes of these students concerning mathematics states of readiness affectively charged?
2. Are the attitudes more or less enduring or characteristic states of readiness?
3. Do the attitudes conform to the normal curve of distribution?
4. What possible implications for teachers of mathematics does this study reveal?

#### I Approach to the problem

- A. Method of measurement: a questionnaire was prepared on a five point scale.
  1. Points on scales: S. A., strongly agree; A., agree; N. O., no opinion; D., disagree; S. D., strongly disagree.
- B. Initial steps in obtaining questions for questionnaire.
  1. Problem was presented to a social psychology class of 30 co-educational students, juniors and seniors, at the Massachusetts State Teachers College, North Adams. (All students had a general mathematics course for one semester in their Freshman year; all had some professional courses in education; all had a course in general psychology. The juniors were also currently taking a course in adolescent psychology; the seniors were currently taking courses in educational psychology and Tests and Measurements.)
  2. Each student prepared a list of fifteen questions which he thought would indicate attitudes toward mathematics, as it is taught in the Junior and Senior High Schools.
  3. A Modified Thurstone-Chave technique was employed. All the questions were listed on a master sheet with a copy available to each member of the class. The class acting as judges

<sup>4.</sup> A Curriculum Guide for Intermediate Grade Teachers, Commonwealth of Massachusetts Bulletin of the Department of Education, 1951, pp. 39-40.

eliminated all obvious duplications, and then selected 130 questions from those remaining which, in the opinion of the majority, would be most likely to measure the attitudes sought. Certain grammatical corrections were made to insure added clarity.

4. A Modified Lickert technique was employed. The 130 questions so derived and arranged on a five point scale were then presented to 65 Sophomores and 78 Freshman at the college. The number of replies to each point on the scale for each question was tallied. The following method was then used to obtain a point score to each question:  $5\sum SA + 4\sum A + 3\sum NO + \sum SD = S$ . The 15 questions with the highest point scores (strongest agreement) and the 15 with the lowest point scores (strongest disagreement) were then selected as the items for a final questionnaire to be given to junior and senior high school pupils.

There are several advantages and disadvantages to this technique. The technique used in preparing the questionnaire insured at least a high degree of objectivity. It seems fairly evident that the questionnaire will not be subjective in the sense that any attitude measured would have been read into the question by the authors. On the other hand, it should be pointed out that the questionnaire has the same limitations as the members of the class who prepared it. For example, they were not mathematics teachers nor were they experts on attitude scales. However, they were closer to the actual school learning situation than more mature teachers.

## II Establishing the scale

The questions were mimeographed, as shown in the following: (Exhibit No. 1). The blank for the name was included because it was believed that if the scale, or a modification of it, were to be used in future clinical work where the student would be identified, any useful interpretations should be made on the same basis. The finished scale was then given in the public secondary schools in the following cities and towns: Adams, grades 10, 11, 12; Ashfield, grades 9, 10, 11, 12; Bourne, grade 9; Charlemont, high school (grades 9-12 but not individually designated); Dalton, grades 7 and 8; North Adams, grades 7, 8, 9, 10, 11, 12; Pittsfield, grades 9, 10, 11, 12; Rockport, grades 9, 10, 11, 12; and Williamstown, grade 11, all in Massachusetts; and Guilford, Conn., grade 7; and Readsboro, Vermont, grade 8.

Each student was instructed to write his name, school, and grade; they were also told the meaning of the points on the scale, S. A., A., etc.; and told that if they were not sure of the question to check N. O. Finally they were told to answer all questions honestly since nothing

## EXHIBIT I

NAME \_\_\_\_\_ SCHOOL \_\_\_\_\_ GRADE \_\_\_\_\_

	S.A.	A.	N.O.	D.	S.D.
1. Learning to use numbers properly will help me in later life.	1.	_____	_____	_____	_____
2. Math is easier for me when my math teacher uses the blackboard to illustrate examples.	2.	_____	_____	_____	_____
3. There should be a special math class for slow students.	3.	_____	_____	_____	_____
4. I think my parents should show me how a family budget works.	4.	_____	_____	_____	_____
5. I would like to have all my math courses presented so as to show relationships to my out of school life.	5.	_____	_____	_____	_____
6. I like mathematics better when the teacher goes very slowly in explaining something new.	6.	_____	_____	_____	_____
7. Girls as well as boys should take mathematics.	7.	_____	_____	_____	_____
8. Arithmetic is helpful to me when I purchase articles at a store	8.	_____	_____	_____	_____
9. I would like to plan a dinner for seven.	9.	_____	_____	_____	_____
10. When the math teacher says, "Are there any questions?", I raise my hand.	10.	_____	_____	_____	_____
11. My teacher seems to like math very well.	11.	_____	_____	_____	_____
12. I should have had better math training in grammar school.	12.	_____	_____	_____	_____

	S. A.    A.    N. O.    D.    S. D.
13. Math is essential in everyday life.	13. _____
14. I can understand a new type of example the first time it is explained.	14. _____
15. We always start a new part of math before I feel sure of the old one.	15. _____
16. High school math is very interesting.	16. _____
17. Mathematics provides for a well rounded high school course of study.	17. _____
18. Arithmetic is useful in the home.	18. _____
19. Math should be required for everyone in junior high.	19. _____
20. I feel math instructors favor superior students.	20. _____
21. Parents should insist that their children spend more time studying math.	21. _____
22. Algebra should be taught in the junior high.	22. _____
23. My parents seem to think that math is easy.	23. _____
24. I am taking math because my friends are taking it.	24. _____
25. I am taking math because my parents want me to.	25. _____
26. When I get a poor mark on a math test I study all the harder.	26. _____

for the next one.

27. Math courses should be practical rather than theoretical.
28. Math problems teach me to think clearly.
29. Most math classes are both pleasant and interesting.
30. Problems in arithmetic teach me to think.

S. A.   A.   N.O.   D.   S.D.

27. \_\_\_\_\_
28. \_\_\_\_\_
29. \_\_\_\_\_
30. \_\_\_\_\_

TABLE I

TOTAL NUMBER OF RESPONSES TO EACH POINT ON THE SCALE  
FOR EACH ITEM

Question	Strongly Agree	Agree	No Opinion	Disagree	Strongly Disagree
1	489	463	17	8	3
2	464	420	74	19	3
3	280	408	172	95	25
4	144	430	266	111	29
5	154	380	336	95	15
6	385	430	89	60	16
7	412	420	107	26	15
8	343	518	68	43	8
9	65	190	410	190	125
10	100	360	312	178	30
11	314	340	380	32	14
12	203	268	248	213	48
13	432	464	55	24	5
14	40	180	280	432	120
15	105	240	192	355	88
16	137	310	310	141	82
17	167	482	265	43	23
18	294	589	58	30	9
19	384	405	112	55	24
20	166	206	278	232	98
21	94	271	282	230	103
22	154	353	237	157	79
23	97	252	345	263	58
24	6	20	154	363	437
25	47	114	206	353	260
26	179	444	210	109	38
27	162	364	393	44	17
28	202	457	201	87	33
29	149	396	184	168	83
30	181	472	205	89	33

TABLE II  
SUMMARY OF MEAN POINT SCORES

Question No.	Total S <sub>t</sub>	7th S <sub>7</sub>	8th S <sub>8</sub>	9th S <sub>9</sub>	10th S <sub>10</sub>	11th S <sub>11</sub>	12th S <sub>12</sub>
1	4.45	4.65	4.38	4.46	4.47	4.41	4.41
2	4.37	4.48	4.23	4.38	4.28	4.36	4.38
3	3.84	3.70	3.82	3.73	3.78	3.85	4.09
4	3.56	3.52	3.48	3.60	3.55	3.48	3.60
5	3.57	3.66	3.75	3.42	3.45	3.58	3.74
6	4.13	4.23	4.16	4.06	4.15	4.23	4.03
7	4.21	4.45	4.26	4.19	4.19	4.20	4.15
8	4.17	4.33	4.15	4.12	4.17	4.17	4.14
9	2.88	3.29	2.95	2.73	2.79	2.90	2.74
10	3.33	3.73	3.68	3.28	3.24	3.23	3.17
11	4.23	4.03	3.67	3.95	4.01	3.84	4.01
12	3.37	3.30	3.23	3.47	3.34	3.56	3.28
13	4.32	4.22	4.42	4.33	4.32	4.30	4.27
14	2.58	3.14	3.11	2.44	2.60	2.39	2.45
15	2.92	2.75	2.47	2.83	2.67	3.00	3.14
16	3.28	3.02	3.11	3.47	3.39	3.11	3.31
17	3.74	3.52	3.61	3.75	3.80	3.76	3.91
18	4.15	4.35	4.23	4.12	4.08	4.14	4.06
19	4.00	4.35	4.20	4.03	4.07	3.85	4.16
20	3.11	3.09	3.09	3.00	2.92	3.13	3.40
21	3.02	3.20	3.48	2.90	2.92	3.05	3.03
22	3.35	2.97	3.23	3.22	3.47	3.37	3.70
23	3.10	3.28	3.30	3.05	3.10	3.05	3.08
24	1.77	1.81	1.99	1.57	1.75	1.88	1.82
25	2.32	2.58	2.66	2.35	2.20	2.77	2.23
26	3.18	4.18	3.81	3.74	3.53	3.35	3.01
27	3.61	3.40	3.49	3.60	3.54	3.74	3.78
28	3.72	3.71	3.96	3.65	3.58	3.62	3.90
29	3.37	3.75	3.72	3.36	3.38	3.15	3.32
30	3.69	3.70	3.90	3.67	3.61	3.59	3.81
M <sub>t</sub>	M <sub>7</sub>	M <sub>8</sub>	M <sub>9</sub>	M <sub>10</sub>	M <sub>11</sub>	M <sub>12</sub>	
3.51	3.61	3.58	3.48	3.48	3.50	3.54	

would be held against them in any way, nor would they identified by name with their teachers. A total of 980 scales were completed.

### III Results

- A. 803 of the questionnaires were scored as follows: an S. A. response was weighted 5; an A. response, 4; N. O., 3; D., 2; and S. D., 1. A total score for each questionnaire was obtained by adding the weighted score for each response. A frequency distribution was then constructed. Ninety was taken the theoretical mean of opinion. The calculated mean was 105.49, with a standard deviation of 9.31. This showed a significant degree of over all agreement for the questions asked. The standard error of the mean was 0.33.
- B. A tally was made of the total number of responses to each point on the scale for each item of the questionnaire. Results are shown in Table I. Similar tallies were made for each grade from 7 through 12.
- C. The following method was used to obtain a scaled score or mean point score for each item of the questionnaire:

$$5\sum SA + 4\sum A + 3\sum NO + 2\sum D + \sum SD$$

N

N equals the total number of responses.

The mean point scores for each item on the scale for each grade and totalled for all grades are shown in the columns of Table II. The mean point score for all thirty questions is shown at the bottom of each column in Table II.

### IV Conclusions

- A. The number of responses or mean point scores to questions 1, 7, 8, 13, 18, 19, indicate that the students' attitudes are affectively charged in relation to the practical value of mathematics.
- B. The number of responses or mean point scores to questions 2, 3, 11, 14, 19, 24, 25, indicate that the students' attitudes are affectively charged in relation to the learning environment as socially conditioned.
- C. The number of responses or mean point scores to questions 2, 3, 6, 14, indicate that the students' attitudes are in agreement with respect to the desire for a more thorough presentation of the subject. The term "always" in question 15 may possibly account for the disagreement in that question.
- D. An analysis of variance of the means by grades gave an F of 4.29, which indicates the differences are not statistically significant, i. e., the lack of change in the attitudes expressed grade by grade

as such, suggests that these attitudes are enduring.

- E. The data in Table I would appear to indicate a close enough approximation to a normal distribution to warrant the conclusion that the attitudes expressed are normally distributed.
- F. The large number of N. O. (no opinion) responses to questions 10, 11, 12, 20, 21, and 23 suggest that students in response to these items hesitate to record themselves in disagreements which imply criticism of parents, teachers, or the school. This further suggests the need for a different technique to determine more clearly these basic attitudes toward the authorities in question.

# A STUDY OF INTERCULTURAL ADJUSTMENT PROBLEMS OF MISSIONARY CHILDREN

HOWARD S. BRETSCH\*  
University of California

PRESENT DAY international and intercultural relationships raise certain questions with regard to youth who must adjust to two or more cultures. Children of foreign diplomats, foreign correspondents, American business employees, representatives of international welfare and service organizations, and missionaries are among this group. Such American youth have a rich opportunity not available to the great majority of American boys and girls. Yet it is also understandable that they may have many problems associated with living in societies with different value systems.

## Purpose

Sons and daughters of missionaries very likely have had to make an adjustment between the cultural pattern of the country in which the mission field was located and the cultural pattern in America. Since the primary purpose of this study was to determine the problems of missionaries' children on coming to America, only the most general inferences may be made for other youth who have been confronted with a similar dual-cultural adjustment problem. The particular conditions that surround life in a mission field (religious in foundation and character), may make it even less appropriate to draw any close parallels between the youth involved in this study and other youth facing the dual-cultural problem. On the other hand the plan of this study may be helpful to those wishing to pursue studies of youth whose dual-cultural problems have had a different setting.

Problems of missionaries' children have not ~~not~~ been recognized. They have been the concern of various mission boards from time to time. Furthermore, an extensive study by Fleming<sup>1</sup>, which in part served as a pattern for the present investigation, revealed that the children of missionaries in India, upon coming to America, could meet new and difficult changes squarely. An earlier report of Yale graduates

\*Assisted by Harold Thomas and Walter Seibert, students.

1. Fleming, Robert L. Adjustment of India Missionaries' Children in America, Ph.D. dissertation, University of Chicago, 1946.

by Huntington and Whitney<sup>2</sup> revealed that the group of missionaries' sons had unusual occupational success.

### Procedure

The sample used for this study consisted of ninety-three missionary children now living in America, many of whom are presently in college. They had lived a major part of their lives in the mission field, had a mean age of 22, and represented mission fields in eight different countries, namely China, India, Brazil, Turkey, Siam, Burma, Korea, and part of Africa. The names of these missionary children were furnished by three Protestant mission boards in America.

A questionnaire submitted to the missionary youth consisted of forty-eight questions involving items of adjustment in the following areas: academic, social-emotional, and social-moral. Nearly all of the questions required two responses, (a) one response applicable to their adjustment in the mission field, (b) one response applicable to their adjustment upon coming to America. With the exception of five questions which called for a free response, the questionnaire could be answered by encircling a choice of one of the appropriate responses applicable to each cultural setting. Most of the items of the questionnaire pertained to middle adolescent years. Since the mean age of the sample was twenty-two, the youth were recalling experiences which they had had but a few years before.

One of the serious problems in a questionnaire is the validation of it. Therefore, care was taken to discuss the questionnaire used in the study with a "trial" group of missionaries' children. This small group which was about fifteen per cent of the final sample was well known to one of the investigators and provided constructive comments for revision of the questionnaire.

### Analysis and Data

In the original study an item analysis was made of the responses to each item and the per cent of youth responding to each was calculated. Although a zero correlation between the various responses to each item may not be assumed to exist since the responses to both phases of each question were made by the same individual, it may be assumed that if a significant difference of the percentages did exist it would be a true difference. The reverse of this need not be so. Therefore, in general, only those differences of percentages which were found to be

2. Huntington and Whitney, The Builders of America, (New York: William Morrow and Co., 1927), p. 368.

significant will be reported in Tables I, II, and III. Citing the other questions which constituted the questionnaire may help to provide some insight into the scope of the study.

Questions not revealing statistically significant differences:

Were you handicapped by being behind in school?

Did you feel successful in your studies?

How were your study habits?

Did you worry about failing in school?

Did you go to teachers for help in problems outside of school?

Did you (a) enjoy, (b) accept, (c) dislike school?

Did you feel self-conscious because of differences in physiological development from that of your peers?

Did you feel superior, on a par, or inferior to your peers?

Did you have the opportunity to read what you wanted?

Were your feelings easily hurt?

Did you feel confident about your judgment in everyday matters?

In what group activities did you engage? (Analysis of the last question by frequency)

Did you wish you were never born?

Did you feel you had an inadequate knowledge of sex?

Was there someone to whom you could tell your troubles?

Did you feel definite aversion to sexual matters?

Did you resent authority of adults?

Did you dislike going to church?

Free response questions:

Comment briefly on your attitude toward social and moral life as you found it in America.

To what extent do you attribute your difficulties in adjustment to the unusual nature of your background?

Briefly, do you feel that you have made a suitable adjustment to modern life in America?

Do you wish to return to the mission field? If you returned to the mission field, it would be because ---

Summary of findings

This summary purports to: (1) summarize the major points as revealed by the data, (2) give an indication of the patterns of response to

TABLE I

PERCENTAGE OF RESPONSE TO PROBLEMS OF ADJUSTMENT BY  
 MISSIONARIES' CHILDREN AND SIGNIFICANT DIFFERENCES BE-  
 TWEEN THE PERCENTAGE OF CHILDREN BOthered BY THE PROB-  
 LEM IN EACH CULTURE

Problem	Columns*				
	I	II	III	IV	III-IV <sup>1</sup>
Were you bothered by class competition?	6	61	25	8	13*
Did you worry about getting "A" grades?	22	62	15	0	15*
Did you feel tense in examination situations?	34	41	18	7	11
Did you want to quit school?	13	63	19	5	14*
Did you feel left out of things socially?	14	41	39	6	33*
Were you worried by a lack of money for future education?	6	71	23	0	23*
Did you feel unskilled in carrying on a conversation?	18	52	22	8	14*
Did you feel you were slow in getting acquainted with people?	25	41	30	4	26*
Did you feel that you lacked a pleasing personality?	14	63	21	2	19*
Were you able to capitalize on some special ability that aided you in gaining social status?	34	35	21	10	11
Did you worry very much?	17	54	27	2	25*
Did you tend to be moody?	29	42	21	8	13
Were you afraid of making mistakes?	36	40	19	5	14*
Did you experience a conflict between your ethical standards and those you found in the world?	32	20	47	1	46*
Did you have difficulty living up to your ethical standards?	20	55	23	2	21*

N = 93

\* Bothered by the problem

I - both in the U. S. and in mission field

II - neither in the U. S. nor the mission field

III - in the U. S. but not in the mission field

IV - in the mission field but not in the U. S.

<sup>1</sup>Differences marked by \*, significant at one per cent level; underlined, significant at five per cent level.

TABLE II

PERCENTAGE OF YOUTH WHO WERE BOthered BY THE PROBLEM IN AMERICA TO THE SAME, MORE, OR LESS DEGREE THAN IT HAD BOthered THEM IN THE MISSION FIELD

Problem	Percentage Bothered			Difference in Percentages More - Less minus
	Same	More	Less	
Did you feel that a lack of money hindered you in making satisfactory social contacts?	59	37	4	33*
Did making friends of the opposite sex bother you? (No response 14 per cent)	18	45	23	22*
Did you feel inhibited by a sense of ignorance of social conventions?	28	60	12	48*
Did you feel that your life was without direction or goal?	61	31	8	23*
Did you daydream about childhood days (No response 3 per cent)	45	37	15	22*
Did you modify your ethical standards?	38	60	2	58*

N = 93

\*Significant at the one per cent level.

TABLE III

PERCENTAGE OF YOUTH WHO WERE BOthered BY SOCIAL CONVENTIONS IN BOTH SOCIETIES AND THE PERCENTAGE OF YOUTH WHO DID AND WHO DID NOT CONFORM TO THE SOCIAL CONVENTIONS UPON COMING TO THE UNITED STATES

Social Convention	U. S.	Mission Field	Differ- ence	Con- formed	Did not Conform	Differ- ence
Dancing	42	22	20*	30	18	12**
Conversation	42	14	28*	27	7	20*
Smoking	39	22	17	20	18	2
Drinking	47	23	24*	20	27	-7
Necking or Petting	52	18	34*	52	22	30*
N = 93						

\* Significant at one per cent level  
\*\*Significant at five per cent level

the questionnaire, omitting any detailed presentation not already shown by the tables.

A review of the tables as well as the data, some of which did not show significant differences, reveals three major points:

1. In response to nearly all of the problems, a greater number of youth were not bothered than were bothered by the problems in the United States and the mission field.
2. Responses to only about half of the items revealed that the problems were significantly greater either in the United States or in the mission field.
3. In every instance where a significant difference existed, the youth were bothered more by the problem in the U.S. than in the mission field.

Since the purpose of this study was to present the problems that missionaries' children had upon making the transition between the mission field and the United States, most of the attention will be given to the problems that seemed to give significantly greater trouble to the youth upon coming to the United States.

In general, the missionary youth in this study fitted fairly well into the academic world without too much stress or strain. The majority had good study habits and did not worry about failing in school in either the mission field or the United States. On the other hand there were more youth who tended to be bothered by various academic problems than who overcame them on coming to America. These problems included: being tense in examination situations, being bothered by class competition, wanting to quit school, and tending to worry over getting "A" grades. The differences in the type of school experiences in the mission field and in America may account for such problems. For many of these youth, school in the mission field consisted, in the early years, of home instruction or attendance at smaller community schools. Over sixty per cent never had any chance to quit school while an additional nineteen per cent who never wanted to quit school in the mission field did want to quit in America. Slightly over half of the missionary youth maintained a healthy outlook on society. Sixty per cent of them stated that the problems of teen age physiological development were not serious. About fifteen per cent felt that both in the mission field and in America they lacked a pleasing personality, were continually left out of things socially and were unskilled in handling social situations. In coming to America about thirty per cent said they took more of an active part in activities and nearly fifty per cent had more contacts with the opposite sex than they had had in the mission field. Gaining of social status was restricted to special abilities by less than a quarter of the group. Having feelings of being left out of things so-

cially upon coming to America was expressed by forty per cent of the youth. Also upon coming to America a lack of money and ignorance of social conventions were problems of about forty and sixty per cent respectively.

With regard to emotional development some problems seemed quite serious. Only a little over half of the youth were able to survive this period of double transition (growing up and cultural acclimitization) without experiencing a lack of confidence in their judgment, without an excess of anxiety, without wishing they had never been born, and lacking the ability to objectify their problems through confiding in others. The transition to the United States was accompanied by the appearance of attitudes and behavior that in themselves were not favorable to good adjustment. While many reported worrying less, one-fifth of the group reported becoming more moody on coming to America.

The area of moral problems seems to suggest that they may have been peculiar to this missionary group and might be less troublesome to other youth having to adjust to two cultural patterns. About sixty per cent of the group said they were bothered about modifying their ethical standards upon coming to America. Forty-seven per cent reported that they experienced conflict among their ethical standards in the United States but did not experience such conflict in the mission field. Twenty-three per cent reported that they had difficulty living up to their ethical standards in the United States, but only two per cent expressed it as being more difficult to live up to their ethical standards in the mission field. In general, social conventions involving moral implications bothered the youth much more in the United States than they did in the mission field.

In summary, data submitted in connection with the free response questions indicated that in coming to America about eighty per cent were shocked by the level of social and moral life here. How much of their attitude can be attributed to the backgrounds of these missionaries' children cannot be determined here. The youths' own feelings indicated that about 50 per cent felt their background had been no handicap in their adjustment while nearly an equal number felt it had been a handicap. About fifteen per cent said they found the culture of the United States new and interesting.

In light of the data presented, practically all of the youth, when asked to respond to the item of general adjustment, felt they had made good or fair adjustments.

Certainly there is not conclusive evidence that these youth have not adjusted reasonably well to two cultures. It can be said, however, that certain problems in the transition from the mission field to America, especially in the areas of social-emotional and moral adjustment, seem to stand out as particularly difficult. Further investigations with other groups of youth faced with dual-cultural adjustments may add to the significance of the findings of this study.

# SOCIO-ECONOMIC STATUS OF RURAL YOUTH BASED ON THE SHORT FORM OF THE SEWELL SCALE\*

WILLIAM H. DREIER  
Iowa State Teachers College  
Cedar Falls

IN A STUDY of the achievement of Minnesota youth attending 6th, 9th, and 12th grades in rural schools the author used the Sewell Socio-Economic Status Scale — short form<sup>1</sup>. The 2,731 pupils who completed this form in the spring of 1948 were attending one room ungraded and larger graded schools located in places with 1940 populations of less than 2,500. The schools included in the study were located in 41 different Minnesota counties. This paper will compare the socio-economic scores made by the rural students who attended the one-room ungraded schools with those who attended the larger graded schools having one or two grades per teacher. It will show the correlation coefficients obtained between the socio-economic scale score and other measures of achievement and ability and will present norms for the Sewell Scale based on this group.

The Sewell Scale was used in the study because a measure of the economic and social level of rural students' homes was needed. Sewell's scale was devised for this purpose.<sup>2</sup> The short form was devised to give essentially the same information as the longer form.<sup>3</sup> The scale had been satisfactorily used by Arny in Minnesota,<sup>4</sup> and by

\*This article is based on data secured by the writer during the spring of 1948 while a graduate student at the University of Minnesota. Suggestions for the collection of the data and the use of the Sewell scale were received from Professors Lowry Nelson and Douglas Marshall of the Rural Sociology Department.

1. Dreier, William H., "The Differential Achievement of Rural Graded and Ungraded School Pupils" Journal of Educational Research, 43:175-186, 1949.

2. Sewell, William H., "The Development of a Socio-metric scale" Sociometry, 5:279-297, 1942.

3. Sewell, William H., "A Short Form of the Farm Family Socio-Economic Status Scale" Rural Sociology, 8:161-169, June 1943.

4. Arny, Clara Brown, Progress Report(1947) on "How Effective is the Home Economics Program in Minnesota Schools", Mimeographed paper. University of Minnesota, St. Paul. 29 p. 1947.

Lundberg and Friedman.<sup>5</sup> More recently Hay reported using the scale in a study of social participation in Maine.<sup>6</sup> The 1950 census of current research projects in Rural Sociology listed a number of studies under way which use the Sewell scale.<sup>7</sup> The widespread and continued use of the scale indicates its usefulness as well as its effectiveness as a measure of the rural family's social and economic status.

#### Description of the Sample

The schools that participated in the study were randomly chosen from all the rural schools in the 41 Minnesota counties. The Fisher and Yates table of numbers was used to select the random group. The proportion of schools selected which participated was 82 per cent for the elementary and 45 per cent for the high school group. The 923 pupils in one or two teacher schools came from 342 schools located in all of Minnesota's eight farming areas outside the St. Paul-Minneapolis urban area.<sup>8</sup> The 599 sixth grade pupils in the one or two grades-per-teacher schools came from 26 graded schools located in six of the eight farming areas of the state. The 1209 ninth and twelfth grade students came from 39 schools located in five different farming areas.

The group of 41 counties in which the elementary schools were located and the 15 counties in which the high schools were located were each compared with the state's 85 rural counties on the following basis:

1. Percentage of farm adults with some high school education in 1940.
  2. Average value of farm products per farm in 1945.
  3. Number of ungraded schools receiving state aid in 1947.
  4. Percentage of 16 and 17 year old farm boys in school in 1940.
  5. Percentage of 16 and 17 year old farm girls in school in 1940.
  6. Culture index based on nationality background, religious composition and intensity of enrollment in elementary parochial schools.
5. Lundberg, George H., and Friedman, Pearl, "A Comparison of Three Measures of Socio-Economic Status" Rural Sociology, 8:227-242, Sept. 1943.
6. Hay, Donald G., "The Social Participation of Households in Selected Rural Communities of the Northeast" Rural Sociology, 15:141-148, 1950.
7. Williams Jr., Robin M., editor "Research Notes" Rural Sociology, 15: 261-268, Sept. 1950.
8. Engene, Selmer A. and Pond, George A., Agricultural Production and Types of Farming in Minnesota, statistical supplement to Bulletin 347 University of Minnesota A.E.S., revised June 1944, 28p.
9. Marshall, Douglas and Peterson, Milo J., Factors Associated With Variations in School Attendance of Minnesota Farm Boys. Bureau of Edu-

The "t" test was used to determine whether the 41 and 15 county means differed significantly from the 85 county population mean on each of the six characteristics. Johnson's formula was used as well as his table for determining the "t" value needed for various levels of significance.<sup>10</sup> None of the values found were significant at the one per cent level.

Additional tests to determine the adequacy of the sample were made of the distribution of the sexes within each of the kinds of schools and of the proportion of high school students who had attended graded and ungraded elementary schools. All the tests indicated the sample had internal consistency and did not differ significantly from the state population.

#### Comparison of Socio-Economic Status of Rural Youth from Small and Large Schools

Are rural boys and girls attending the small one-teacher schools in a different socio-economic level from that of pupils who attend a larger school where there is a teacher per grade? The Sewell Scale scores made by the rural youth in the 6th, 9th, and 12th grades of the study were compared on the basis of mean scores of pupils from the two kinds of schools. The results of these comparisons are shown in Table I.

The sixth grade pupils were attending either an ungraded (one or two teacher school) or a graded (one or two grade-per-teacher school) at the time they were tested. The ninth and twelfth grade students were attending high school at the time they completed the status scale. Information from a questionnaire completed by these same students made it possible to determine the kind of elementary schools the students had attended. Each student's card was coded according to the years of education received in the ungraded and graded elementary schools. Only the cards of those 6th, 9th and 12th grade students who had attended each kind of school (ungraded or graded) for half a year at their elementary education were included in the data presented in Table I.

Table I indicates a general tendency for the students from the small rural school in each grade to have a slightly lower mean Sewell Status Scale score than those who attended the larger school. The standard deviations differ very slightly with a general tendency for the deviations to be smaller as the grade level increases. The "t" test suggested by Johnson<sup>11</sup> was used to determine whether the greatest differences were

cational Research, Graduate School and the Agricultural Experiment Station, University of Minnesota. July, 1948. 24p.

10. Johnson, Palmer O., Statistical Methods in Research. Prentice-Hall, Inc. New York, 1949. 377p.

11. Johnson, op. cit. p. 72

TABLE I

SEWELL SOCIO-ECONOMIC SCALE SCORES OF MINNESOTA RURAL  
YOUTH ATTENDING 6th, 9th AND 12th GRADES OF ONE-TEACHER  
AND LARGER RURAL SCHOOLS

Grade	N	Mean	Standard Deviation
<u>One or two teachers per school - Ungraded</u>			
6	651	65.76	12.54
9	206	66.46	12.09
12	159	68.13	10.77
Total	1016	66.27	12.23
<u>One or two grades per teacher - Graded</u>			
6	229	67.27	12.57
9	162	67.31	13.08
12	91	70.46	10.41
Total	482	67.89	12.41
<u>Above groups combined</u>			
6	880	66.15	12.58
9	368	66.83	12.56
12	250	68.98	10.68
<u>Grand total</u>	1498	66.79	12.32

significant. The standard deviations of the 6th and 12th grade youth in the ungraded, graded and combined groups were compared. None of the differences between 6th and 12th grade rural youth were significant.

The results from these comparisons suggest that the rural youth from the graded and ungraded schools at these grade levels do not differ significantly in their Sewell Socio-Economic Status Scale scores. The total group of 1498 students would seem to be homogeneous as far as their socio-economic status is concerned as measured by the scale.

#### Correlation of Socio-Economic Status with Measures of Intelligence and Achievement

A socio-economic status scale should not measure the same thing measured by tests of mental ability or academic achievement. The extent to which measures correlate with one another indicates in some way the extent to which each is measuring the same thing. The Sewell Scale was correlated with the Otis Gamma test of mental ability and with three of the basic skills measured by the Progressive Achievement Tests.

The correlations between the Sewell Socio-Economic Scale scores and the Otis IQ, reading, language and arithmetic achievement test scores are very low and sometimes negative. But most often they are positive and have large standard errors. In all of the 22 comparisons only two of the correlations are significant. These correlations are between the Sewell Socio-Economic score and the Otis IQ. They amount to .40 and .20 and are based on two groups of 55 and 221 12th grade students from graded elementary schools. It may be concluded that the Sewell Socio-Economic Status Scale score does not indicate or predict the test scores of reading, language, arithmetic, or mental ability.

#### Percentile Norms for Sewell Socio-Economic Status Scale Based on Rural Youth in School

The short form of the Sewell Socio-Economic Status Scale was devised to secure information about farm families which could be scored to indicate the relative status of the family. This score is meaningful only when it can be compared with the scores of others. The uniformity of farm living throughout the Mid-west suggests that scores collected from a large number of counties located in different farming areas in one of the states of the corn belt could serve as norms for scores collected in other mid-west corn belt states. In order to make raw scores more meaningful a percentile score equal to a raw score of the Sewell Socio-Economic Status Scale is presented in Table II.

TABLE II

PERCENTILE NORMS FOR THE SEWELL SOCIO-ECONOMIC STATUS SCALE (SHORT FORM) BASED ON 1948 MINNESOTA RURAL YOUTH IN THE 6th, 9th AND 12th GRADE

Mean = 66.79			Standard Deviation = 12.32		
Percentile	Raw Score	Percentile	Raw Score	Percentile	Raw Score
98	91	69	73	24	58
96	89	66	72	21	57
94	86	62	70	18	56
92	84	58	69	16	54
90	83	54	68	14	53
88	82	50	67	12	52
86	80	46	66	10	51
84	79	42	64	8	50
82	78	38	63	6	47
79	77	34	62	4	45
76	75	31	61		
73	74	27	59		

Conclusions and Summary

This paper had indicated that the 1498 rural youth who form the norm group for the percentile scores of Table II come from all eight of the farming areas in Minnesota as well as from the major kinds of schools in that state. The mean Sewell scores of the 6th, 9th and 12th grade groups vary but are not significantly different. The total group is homogeneous as far as its socio-economic status is concerned. The Sewell scores were correlated with measures of achievement and mental ability and in general were found to have little, if any, value in predicting these measures. Future users of the Sewell Socio-Economic Status Scale may find their scores more useful and meaningful when compared with the percentile norms established by this sample of Minnesota rural youth.





# RESEARCH PRIORITIES IN EDUCATIONAL ADMINISTRATION

A. P. COLADARCI, E. BROOKS  
and W. R. ODELL  
CPEA Administration Center  
Stanford University

IN RECENT literature, professional meetings and informal discussions, the school administrator has been reflecting a concern that educational research activities are not sufficiently germane to the problems confronting him. From that perspective, the researcher has been pursuing interesting but largely "theoretical" problems and avoiding or indicating an insensitivity to the many "practical" and immediate unresolved indeterminate situations that daily confront those who are engaged in organizing and maintaining the ongoing educational establishment. Although the extent and detailed nature of this dissatisfaction remain to be determined empirically, one can be reasonably safe in the assumption that the situation exists and is serious enough to warrant examination. For instance, despite some methodological questions that arise in connection with the survey by Johnson (3), one cannot fail to be concerned that "only 35 percent of the administrators responding were satisfied with what research workers are contributing to help them improve their school programs . . . and . . . according to 77 per cent of those replying, research reports are too formal and foreign for practical application in ordinary school situations." (3:8).

It is felt here that the criticism referred to has not been answered adequately by educational research people. To insist, with justification, that the "practical" and the "theoretical" constitute an integrity, is not convincing if the researcher who so insists continues to be exclusively theoretical in his discourse and operation. To insist that the "practical" must be deduced from the "theoretical" is not helpful if the research person does not assist in such deduction and hypothesizing. Furthermore, it does not seem presumptuous to ask that the educational research person be interested continuously in how his data, hypotheses and theoretical systems bear upon the very practice that he presumably is attempting to clarify and systematize. Nor is it indefensible to suggest that the educational researcher should be sensitive to and should participate in attempts to clarify the indeterminate situations confronted by those who are using the results of his research in educational practice.\*

\*We are mindful of a corollary of these propositions; the administrator, similarly, should have the research sophistication necessary to operate effectively with researchers and research data. The criticism that research people and data are not helpful would be offered less frequently.

If the educational researcher is to be encouraged to address himself to problems considered urgent by school administrators, it is relevant to raise the question of what are these problems or problem areas in the context of present educational administration? The question appears to be unanswered at the moment. This is not to gainsay the heuristic value of Moore's study (4), Reller's excellent evaluative summary and analysis (5), or Hunt's "sample questions" (2). However, the first is a local investigation and whether the latters' hypothesized problems (5:9) (2:56-57) are those of present administrators and, if they are, the relative priorities assigned to each by such administrators are still matters for conjecture.

The inquiry reported below was undertaken for the purpose of providing some clarification of the question of research priorities from the perspective of the large city superintendent. \*\* The superintendents of the five largest cities in each of the forty-eight states constituted the panel and were identified through the listings in the most recent AASA Yearbook (1). Each of the 240 individuals was asked, by letter, to respond to the following question: "As you look over the contemporary unresolved administrative problems faced by you in discharging the responsibilities of a city school superintendent, which one, or ones, do you see as most crucial and deserving research priority?"

One hundred and sixty-nine usable responses (70.4%) were received from forty-six states. The average number of research problems submitted by each superintendent was 3.2 with a total of 541 items. The problems suggested were pitched at varying levels of specificity—some very generally stated ("What is the responsibility of the school in a democratic society?") and others quite specific ("How do other city superintendents budget their time?"). Such heterogeneity in levels of discourse makes for difficulty in an attempt at classification and it is quite apparent in retrospect that other analysts might obtain equally convenient but different categories and at different levels of abstraction. To prevent the exclusion of all, or nearly all, of the replies, we chose to identify broader categories than might be desirable in an inquiry of this kind. It was felt that the analysis would retain considerable clarification value nonetheless, pointing to areas of perceived urgency and identifying bases for detailed investigation of the basic question.

perhaps, if administrators were trained to use research findings more often and more adequately.

\*\*This administrative level and context was chosen rather arbitrarily not on the assumption that large city superintendents are the most competent judges of research priorities. Rather, it was felt that some homogeneity of role and responsibility must obtain among the panelists and it appeared that the "large city superintendent", as defined, reasonably satisfied such a criticism.

The rubrics abstracted from the superintendents' responses are presented below, the parenthetical figure following each rubric indicating the number of the 541 "research priority" problems falling under that heading. It will be noted that the major (upper case) headings account for all of the 541 problems submitted, as do the subheadings (*italicized*). The specific problem areas listed under the subheadings, however, are only those that could be assigned frequencies of more than 3. These more specific areas, therefore, do not account for all of the problems submitted — the relatively "unique" ones having been omitted from this analysis.

#### The School Staff (192)

##### The Teaching Staff (75)

- Improvement of recruitment and selection (27)
- Improvement of teacher evaluation and selective retention (22)
- Improvement of in-service training (15)
- Methods for obtaining and maintaining good morale (10)

##### The Superintendent (53)

- Methods for budgeting personal time adequately (15)
- Definition and use of democratic procedures (13)
- Methods for delegating responsibility (9)

##### The Professional Administrative Staff (42)

- Definition of duties and responsibilities (21)
- Determination of optimum size (8)
- Improvement of in-service training (7)
- Methods for evaluation of staff (5)

##### The Supervisory Staff (18)

- Methods for organization (7)
- Improvement of supervision methods (6)
- Supervisor selection and evaluation (4)



#### The Clerical -Secretarial Staff (4)

#### The Educational Program (123)

##### Curriculum (75)

- Definition of basic principles of and effective procedures for general curriculum development (40)
- Improvement of curriculum for the slow learner (13)
- Relationships between curriculum and holding power (5)
- Improvement of health education (4)

Instruction (29)

- Methods for validating present procedures (15)
- Adapting instructional methods to class size (6)

Guidance (19)

- Methods for promoting more effective inter-group and inter-individual behavior (9)
- Development of improved bases and procedures for grouping (7)

Public Relations (110)Community Relations (86)

- Improvement of general public relations program (40)
- Methods for encouraging and using lay participation (15)
- Methods for handling community pressure groups (15)
- Methods for reacting to "attacks on the schools" (9)

Relations with School Agencies and Professional Groups (24)

- Definition of optimum relationship between local school board and school staff (15)
- Methods for maintaining relations with professional organizations in education (4)

Plant Planning and School Finance (88)School Finance (57)

- Methods for obtaining increased financial support (45)
- Improvement of budgeting and accounting procedures (10)

Plant Planning (31)

- Definition of optimum plant size (9)
- Estimation of building costs (7)
- Determination and use of population trends (7)
- Reduction of building costs (6)

Role and Responsibility of the American Public School System (24)Unclassified (4)

Several things are apparent in the foregoing data. First, it seems clear that the nation's large city superintendents are not in agreement with regard to what constitute priority research problems in their operations. Of the specific problem areas used above, it will be noted that the one accounting for the largest number of submitted problems is "Methods for obtaining increased financial support" — and in that case the frequency was only 45 (8%). Such a state of affairs is not necessarily an undesirable one nor does it attenuate the force of the administrators' criticisms and suggestions. Perceived problems are directly related to perceived purposes and interests. School

systems and school people differ, more or less, in both regards. When a wide geographical area is involved and when the respondents are those responsible for the total educative process in their areas, one might expect considerable heterogeneity in purpose and interest and a corresponding "disagreement" with the question of what are important research problems.

It is equally patent that the perceived problems would involve almost all of the present research techniques and interests for their resolution. The data do not support the rather strong notion that administrators are interested merely in descriptive-survey research problems. An education researcher, whatever his particular methodological competencies and research persuasion may be, will find problems to his liking. Furthermore, the substantive nature of the administrators' concerns is a most catholic one. No preoccupation with a particular aspect of the total educative structure and process is seen here. When the administrator is critical of educational research activities he evidently does not have a particular substantive area or educational interest in mind. His concern appears to embrace all dimensions of the educative process and organization and, again, the particular substantive interest of an educational research person is not a limiting factor in his ability to meet the research needs described by the respondents.

The research person in education will probably find in these data a basis for negative criticism. He might venture the observation that many of the problem areas identified by the respondents are areas in which much research has already been done and usable data are available — i. e., we have answers for some of the problems. If such is the case, either the findings have been deemed inadequate by the respondents or the findings are unknown to them. Both possibilities undoubtedly obtain in large measure and, in part, merely reassert the administrators' criticisms of the researcher, on one hand, and suggest the need for some research training for the administrator on the other. If research findings are not being communicated to those who can use them, the fault is as much with those researchers who do not communicate adequately under present circumstances as it is with those administrators who are disinterested in or unprepared to join in such communication.

The need for more detailed inquiry into this situation and systematic discussion of it is evident. The nature and extent of cooperation received in the present survey indicate that at least these administrators are both vocal and interested. If educational research people are similarly concerned and interested (and we believe that they are), joint attempts to clarify and improve the relationships between research activities and administrators' concerns undoubtedly will yield a profit for both groups.

## BIBLIOGRAPHY

1. American Association of School Administrators, American School Curriculum, Thirty-First Yearbook, 1953. Pp. 361-542.
2. Hunt, Harold C., "Some Questions Educators Ask," Phi Delta Kappan, 35:56-58, 68, October, 1953.
3. Johnson, Loaz W., "What Administrators Want and Will Use from Research Workers," American Educational Research Association 1949 Official Report, pp. 7-12.
4. Moore, Hollis A., "Blind Spots in In-Service Education for Administrators," Nation's Schools, 51:43-46, April, 1953.
5. Reller, Theodore L., "Administration - City School," in W. S. Monroe (ed.), The Encyclopedia of Educational Research, New York: Macmillan, 1950. Pp. 7-18.



# TEACHERS' EVALUATIONS OF THE MENTAL HEALTH STATUS OF THEIR PUPILS\*

GWEN ANDREW

Michigan Dept. of Mental Health

HILDA LOCKWOOD

Battle Creek Public Schools

OFTEN IN developing a program to best meet the needs of our school children we neglect to fully consider the present situation. Without such consideration any change in approach may lose much of its effectiveness and become mere floundering, with the rather aimless goal of trying to prove to ourselves, at least, that we are "keeping up." Therefore, a most important part of any school program is an occasional systematic survey of the status quo. The information offered through such a survey may then be incorporated into an over-all plan for the school program with much hope for an effective change which will benefit the children involved.

In Battle Creek the school program is constantly under consideration by the Council on Instruction, which is made up of a teacher representative from each of the seventeen schools in the system and the Director of Curriculum. This council, with the above stated assumption in mind, became interested in studying the health program of the schools and community in relation to its influence on the children. Four committees were appointed from the council to study the following aspects of the health program: (1) the instructional program as it pertains to health, (2) available health services, (3) health status of the children at the present time, and (4) the community as it impinges on the health of the children. As a part of the work of the health status committee, representatives of the school group and the Michigan Department of Mental Health planned and carried ~~each~~ <sup>out</sup> ~~of~~ <sup>the</sup> survey of the mental health of the children enrolled in the schools. It is this survey with which this paper concerns itself.

Simply stated, the method of the study was to have the teachers rate a randomly selected sample of ten per cent of the 8,500 children enrolled in the schools on an eleven-item rating scale of pupil adjustment. This rating scale, constructed in mental hygiene terms, attempted to evaluate each child on the basis of his observable behavior manifested in the classroom. These eleven items, each having a five-point scale, included the following:

\*This article is based on a Michigan Department of Mental Health report entitled "A Survey of the Mental Health Status of Battle Creek Public School Children."

1. Over-all Emotional Adjustment — total emotional adequacy in meeting the daily problems of living as manifested in school.
2. Social Maturity — ability to deal with social responsibilities in school, in the community, and at home, appropriate to the pupil's age.
3. Tendency Toward Depression — tendency toward melancholia, depressed mood reactions.
4. Tendency Toward Aggressive Behavior — overt evidence of hostility and/or aggression toward other children and/or adults.
5. Extroversion-Introversion — tendency toward living outwardly and expressing his emotions spontaneously versus tendency toward living inwardly and keeping emotions to himself.
6. Emotional Security — feeling of being accepted by and friendly toward one's environment.
7. Motor Control and Stability — capacity for effective coordination and control of motor activity of the entire body.
8. Impulsiveness — tendency toward sudden or marked changes in mood.
9. Emotional Irritability — tendency to become angry, irritated or upset.
10. School Achievement — over-all evaluation of the pupil's competency in school subjects relative to his own age group.
11. School Conduct — conduct in the classroom situation as evidence of his ability to accept the rules and regulations of the school community.

In addition to the rating scale, background data was requested on each child which included the following: birthdate, sex, father's occupation, race, intelligence test scores, the parents or others with whom the child lives, and the number, relative age and sex of his siblings. The father's occupation was used as an index of the socio-economic level of the home, and the relative age of the siblings made it possible to study the relationship between a child's adjustment and his ordinal position in the family as well as the relationship between adjustment and the number of siblings.

Once the children were rated they were considered as well or poorly adjusted on the basis of a composite score computed from the rating scales. This immediately gave an indication of the number of maladjusted children in the schools. Following this the background data pertaining to each child was compared with his adjustment rating and some interesting and informative factors emerged.

1. Eighty-one per cent of the children were found well adjusted, and 19 per cent poorly adjusted. Two per cent were found severely maladjusted. In terms of the total school population of 8,500 we would expect to find approximately 1,615 poorly adjusted children, but only approximately 170 of these poorly adjusted children would be severely

maladjusted.

2. There are no differences between grades in the per cent of well or poorly adjusted children included within them.
3. On the sample surveyed there was some relationship between poor adjustment and low socio-economic status.
4. Some relationship was found between poor adjustment and low intelligence. This should not be interpreted to mean that poor adjustment did not exist at higher intelligence levels.
5. There was a relationship between poor adjustment and the parent figures with whom a child lived. That is, a large percentage of poorly adjusted children did not live with both natural parents.
6. There was no relationship between the child's ordinal position in the family and his adjustment.
7. No relationship was found between the child's adjustment and the number of children in his family.
8. No significant relationship was found between a child's race and his adjustment.

These factors or results answer many questions about the present mental health status of Battle Creek school children and they indicate several problems to be considered in future planning. First, they tell us the approximate amount of maladjustment present in the total school population. Secondly, they tell us some factors related to poor adjustment which the teacher, if she is aware of their existence, may use as warnings to be alert for signs of emotional problems in the child involved.

There is little that the schools themselves can actively do about many of the factors found to be related to adjustment. For example, the school system can hardly be responsible for improving the socio-economic status of the families of its children beyond the intrinsic potentiality which may be developed by a sound education which will be of value to those children when they become adults and parents of future generations. Nor can the schools be expected to solve the problems of the home which is broken through death or divorce.

In spite of this, there are important steps the school system may take to improve the adjustment of its pupils, and it is with these that the Battle Creek Public Schools are concerned. Those children who are poorly adjusted need various kinds of help and much of it might come through the schools. Many of these children can be helped indirectly through the general efforts of the school to apply mental health principles to its curriculum.

The severely maladjusted children may need concentrated psychotherapy and they should have it as soon as possible. This calls for a staff of professionally trained persons. Conceivably, such a staff might be a part of the school system, but there are many practical considerations involved. Chief among these are the expense of maintaining

such a staff and the effects on the therapeutic program of the necessarily authoritative connotation given such a clinic because of its relationship with the administration of the schools. However, there may be many advantages to a clinic within the schools. Primarily, this would ensure the availability of clinical help and it might have the tendency to dispel the too prevalent feeling that it is unusual or strange to make use of such help.

Many states, including Michigan, now operate child guidance clinics with teams of professionally trained persons. The clinical services of these persons are available for referrals from schools. The visiting teacher is often the liaison between schools and clinics. Where these services are available they should be utilized promptly for those children whose problems involve serious emotional disturbances.

Perhaps a more immediate problem for practical solution is that of what can be done within the entire curriculum to help all children toward good adjustment. First, it seems reasonable to assume that no curriculum can be geared to develop and maintain good mental health in school children without a continuing program of in-service training for the teachers who work with these children. This program may best be carried out through workshops and courses conducted with the help of those professionally trained in the area of mental health. For the past several years, Battle Creek schools have been conducting such workshops for the teachers.

Most mental health principles will necessarily need to be applied in each class to each group of children and to each individual, rather than taught in special classes. However, some mental hygiene classes with carefully planned content might also be offered, particularly to high school students. Frequently, lectures from an outside "expert" have been used in high schools and this practice may be the best method, since a school system may not have a person prepared to carry out the function. A school fortunate enough to have a clinically trained person as part of the staff might well use this person for such lectures.

The media of group and extracurricular activities have often been stressed and it is through these media that much could be done to give satisfaction to those children who have been indicated as not seriously maladjusted. The effectiveness of these group activities will depend upon the skill of well-trained group leaders and this skill might well be developed through the teacher-training mentioned previously. The actual method of approach must have careful consideration before it is put into action.

Further study must be carried out at regular intervals to determine the effectiveness of whatever methods are used as preventive measures, or as methods to improve the mental health status of the poorly adjusted school children. Without this study the schools cannot know the scope of the problems of the children at any specific time, nor

can the effectiveness of the efforts they are making to help children develop and maintain a good adjustment be evaluated.





# RELATIONSHIP OF PRIMARY ABILITIES SCORES AND OCCUPATIONAL CHOICES

LOUIS G. SCHMIDT, Eastern Illinois State College,  
Charleston, Illinois  
J. W. M. ROTHNEY, University of Wisconsin

TO WHAT EXTENT are one's test performances related to the type of occupation he chooses when he is a sophomore in high school? The answer to this question is important from the standpoint of its bearing on the theory of mental measurement and its implications for vocational and educational guidance. Guidance workers long have looked for instruments that would point quickly to fitness for specific occupations but up to the present time their search has not revealed any which indicate that a pupil is cut out for a specific career and that in no other will he find equal success and satisfaction. Yet the idea that it can be done has been fostered by many testing agencies with unpleasant consequences for the individual who sets for himself a too restricted or too ambitious goal. Currently it seems that an individual can find stimulation and satisfaction in more than one specified job.

If one reads the manuals of most tests very carefully he will usually find that test authors have urged caution in the use of test scores or have suggested that they be supplemented by data from other sources. Unfortunately some authors omit such cautions and some users of tests overlook those that are presented. As a result the implication is often drawn that certain measured performances are essential if the person is to succeed in a given occupation and that choices which go contrary to the measurements are likely to be unwise. In the following pages a report of relationships between tested performances and occupational choices is presented.

The test of Primary Mental Abilities published by Science Research Associates, was selected for this investigation. Ability of the reputations of its authors and the fact that it provides five ability scores which the test manual indicates are necessary for success in selected occupations. A high score on the space factor, for example, purports to indicate the desirability of certain educational and occupational choices.

"Space (S) is the ability to think about objects in two or three dimensions. Blueprint reading, for example, requires this ability. The designer, electrician, machinist, pilot, engineer, and carpenter are typical workers who need ability to visualize objects in space. S is helpful in geometry, mechanical

drawing, art, manual training, physics and geography classes.\*

If the above statement holds and the same principles apply in the other factors the tests would indeed be valuable aids to counselors. In this study we set out to find whether students who chose various occupations when they were high school sophomores actually scored high in the factors said to be necessary to perform that occupation.

Nine hundred students, the entire tenth grade classes of each of the four Wisconsin high schools used in the Wisconsin Counseling Study were selected as subjects of this study. The average chronological age of the experimental group was fifteen years and five months on September 1, 1948. Four hundred and fifty-six were girls and 444 were boys.

Career choices of the subjects were obtained at the time each subject took the Henmon-Nelson Test in October 1948, again during a thirty minute interview with each subject, and finally after each subject took the Primary Mental Abilities Test in April 1949. Stability of the choices was determined by comparison of the three that were made during the year. They were considered stable if all three were identical or if the one given during the interview was identical with one of the other two. All other combinations of choices were labeled "no choice" for purposes of this experiment. Eighty-three per cent of the experimental group were found to have stable occupational choices according to these criteria.

The ten occupations chosen most frequently by the students were secretary, teacher, nurse, artist, machinist, auto mechanic, doctor, draftsman, engineer, and bookkeeper. Although farmer was the third most popular selection, it was dropped from consideration in this study because of the wide variety of jobs and skills represented in that occupation.

To obtain a consensus of estimates about the importance of measured performances in each of the ten selected occupations, four sources were used. The first was a study of the occupations listed under each ability in the Primary Mental Abilities manual; the second was an analysis of performances listed in the United States Employment Service General Aptitude Test Battery; a third was an examination of job analyses made by the United States War Manpower Commission for each of the ten occupations; and the fourth was obtained from the "Abilities Required" section of Occupational Abstracts, published by New York University.

Results of the study were obtained from the application of three basic techniques. The first procedure was that of comparing Primary

\* Thurstone, L.L., Manual of Instructions for the Primary Mental Abilities Test (Chicago: Science Research Associates, 1947).

Mental Abilities subtest scores of subjects who chose one of the ten most common occupations, with the scores said to be important for success in their occupational choices. The data were next analyzed to determine the relationship of the Henmon-Nelson Test of Mental Ability scores to the occupational choices of subjects who had achieved significantly high or low scores on individual subtests on the Primary Mental Abilities Test. Finally, a study was made of the tenability of the hypothesis that subjects tended to choose occupations which were said to require the test proficiencies they had achieved.

Study of the significant differences between scores made by subjects who selected certain occupations, and the scores of those who did not, reveals that:

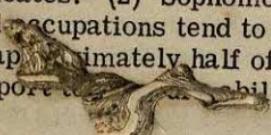
1. Subjects selecting secretarial work as an occupation showed significant superiority in the Verbal and Reasoning subtests, but not on the Word-Fluency subtest. Examination of the literature in the sources given above suggested that Verbal Reasoning, and Word Fluency were desirable qualifications for this occupation.
2. Subjects selecting teaching as an occupation made significantly higher than average scores on the Verbal and Reasoning subtests. As above, Reasoning, and Word Fluency had been said to be important for this occupation.
3. Subjects selecting nursing as an occupation made significantly higher than average scores on only the Reasoning factor. Word fluency had also been indicated as desirable on the criteria given above.
4. Subjects selecting art as an occupation did not make significantly higher than average scores on the Space subtest.
5. Spatial and Number Ability are described in the test manual as being important for machinists, but the subjects who selected machinist as an occupation failed to make scores significantly different from the average of all other subjects on either of these subtests.
6. The same conditions as in (5) above applied to subjects selecting auto mechanic as an occupation.
7. Subjects selecting doctor as an occupation failed to make significantly high scores on the Space and Reasoning subtests that were said to measure abilities necessary for doctors.

8. Subjects selecting drafting as an occupation showed superiority on the Space subtest, but not on the Number and Reasoning subtests. All three had been said to be essential abilities for draftsmen.
9. Spatial and Number abilities were said to be desirable for engineers, but subjects selecting engineering as an occupation failed to make significantly higher than average scores on those subtests.
10. Subjects selecting bookkeeping as an occupation made significantly high scores on the Number subtest but not on the Reasoning subtest. Both were said to be desirable abilities for bookkeepers.

A summary of the results presented above indicates that in five of the ten occupations considered, the subjects made significantly higher than average scores on subtests purporting to measure abilities important for success in the occupations.

The coefficient of correlation between the total raw scores on the Primary Mental Abilities Test and the Henmon-Nelson Test of Mental Ability with a population of 900 was found to be .38. There is, therefore, some evidence that the total scores on the Primary Mental Abilities test provide data similar to those provided by scores on the Henmon-Nelson Test of Mental Ability. The coefficient is not high enough, however, to make accurate prediction of an individual's performance on one of these tests from knowledge of the other.

It is concluded that: (1) Occupational choices of tenth grade secondary students are more stable than previous literature in this area indicates. (2) Sophomores in high schools who select certain specified occupations tend to make significantly higher than average scores on approximately half of the tests of Primary Mental Abilities which purport to measure abilities required for those occupations.



# JOURNAL *of* EDUCATIONAL RESEARCH

---

Volume XLVII

May, 1954

Number 9

---

## PREDICTING PERFORMANCE IN THE TEACHING PROFESSION

A. B. CARLILE  
Butler University  
Indianapolis, Indiana

Purpose. Specifically, this investigation was made for the purpose of answering the question: "What relationship exists between the grades of students in student teaching and the scores which they made on selected measures in the following fields: (1)intelligence, (2) teaching aptitude, (3) scholastic achievement, (4) proficiency in the basic skills and (5) personality traits as they are revealed by personality inventories and personality schedules?"

Source of Data. Grades in student teaching were copied from the Student Record sheets which are on file in the office of the Registrar of Butler University, Indianapolis, Indiana. Measurements in the five fields named above, were obtained as indicated below:

1. Intelligence  
Detroit Advanced Intelligence Test  
Henmon-Nelson Tests of Mental Ability for College Students.
2. Teaching Aptitude  
Cox-Orleans Prognosis Test of Teacher Ability  
Ohio State University Psychological Examination  
Estimates of the Faculty of Butler University
3. Scholastic Achievement  
Grade Points in College Work  
Columbia Research Bureau American History Test  
Columbia Research Bureau English Test  
Columbia Research Bureau Civics and Government Test
4. Proficiency in the Basic Skills  
Iowa Every-Pupil Test in Reading  
Iowa Every-Pupil Test in Work-Study  
Iowa Every-Pupil Test in Language  
Iowa Every-Pupil Test in Arithmetic

### 5. Personality Traits

Allport A-S Reaction Study

Bell Adjustment Inventory

Bernreuter Personality Inventory

The following numerical values were used in making statistical computations: (1) raw scores obtained on tests; (2) faculty estimates, (Percentile Ranks) obtained on a rating scale with values from zero to one hundred; (3) grade points per hour of credit and (4) the values three, two, and one, which were assigned to the grades A, B, and C, respectively. The findings are given under the following headings:

Intelligence. The relationship found between grades in student teaching and intelligence test scores is shown in Tables I and II. The frequencies in both tables reveals a tendency for the high intelligence scores. Further, the coefficients of correlation are positive. The correlation with the Detroit Intelligence Test is statistically significant but low. Its forecasting efficiency is four per cent. The correlation with the scores on the Henmon-Nelson Tests of Mental Ability is too low to be significant.

Teaching Aptitude. Three measures, the Ohio State University Psychological Examination, the Cox-Orleans Prognosis Test of Teaching Ability and Faculty Estimates, were used to measure teaching aptitude. The relationship found between each of these and grades in student teaching is shown in Tables III, IV and V. The scatter of the scores on this test indicate a positive relationship. The correlation coefficient, however, shows that this relationship is very low.

The frequencies of the Cox-Orleans Test indicates a low correlation. The individuals making the highest scores on this test did not receive grades of "A" but grades of "B" in student teaching. Similarly, the lowest score was not recorded for a student who received a grade of "C" but for one who made a grade of "B". The means of the scores on this test do not decrease as the grades decrease. The mean is highest for the students who made grades of "B". The coefficient of correlation is too low to be significant.

The scatter of the frequencies of faculty estimates indicates a lack of relationship but the means vary in the same direction as the grades. The coefficient of correlation is positive but indicates a relationship so low that it is almost without value as a means of forecasting grades in student teaching.

Scholastic Achievement. The relationship found between grades in student teaching and the measures of scholastic achievement may be noted in Tables VI, VII, VIII and IX. There is a fairly high positive

TABLE I

THE STUDENT TEACHING GRADES OF EIGHTY-THREE STUDENTS AND THEIR SCORES ON THE DETROIT ADVANCED INTELLIGENCE TEST

Detroit Intelligence Test	Student Teaching			Total f
	A	B	C	
230-239	1			1
220-229	1	2		3
210-219	1	1		2
200-209	1	3		4
190-199	2	4	1	7
180-189	2	3		5
170-179	3	4	3	10
160-169	2	11	1	14
150-159	6	1	4	11
140-149	2	7	3	12
130-139	1	6	1	8
120-129	1		1	2
110-119		1	1	2
100-109			1	2
90-99		1	1	2
				
Totals	23	44	16	83
M (Detroit)	173	166	151	165
r=.28± .068				

TABLE II

THE STUDENT TEACHING GRADES OF NINETY-  
 SEVEN STUDENTS AND THEIR SCORES ON THE  
 HENMON-NELSON TESTS OF MENTAL ABILITY  
 FOR COLLEGE STUDENTS

Henmon-Nelson Mental Ability	Student Teaching			Total <i>f</i>
	A	B	C	
77-79		1		1
74-76	2			2
71-73	1	2		3
68-70		4	1	5
65-67	1	2		3
62-64	1	2	1	4
59-61	4	9		13
56-58	2	2	3	7
53-55	3	4		7
50-52	4	2	1	7
47-49	1	6	1	8
44-46	1	7	3	11
41-43		5	3	8
38-40	3	3		6
35-37	1	1	1	3
32-34	1	1		2
29-31			1	1
26-28		3	1	4
23-25		1	1	2
Totals	25	55	17	97
M(Hen.-Nel)	55	52	46	52
r=.23 ± .064				

TABLE III

THE STUDENT TEACHING GRADES OF NINETY-EIGHT STUDENTS AND THEIR SCORES ON THE O. S. U. PSYCHOLOGICAL EXAMINATION, FORM 20

O. S. U. Examination	Student Teaching			Total <i>f</i>
	A	B	C	
140-149	1			1
130-139	2	3		5
120-129	2	4		6
110-119	3	2	2	7
100-109	2	3		5
90-99	4	5	2	11
80-89	2	7	4	13
70-79	1	5	1	7
60-69	2	8	1	11
50-59	1	6	2	9
40-49	1	5	4	10
30-39	2	6	2	10
20-29		2	1	3
Totals	23	56	19	98
M (O.S.U.)	93	76	68	78
r=.12±.06				

TABLE IV

THE STUDENT TEACHING GRADES OF SEVENTY-NINE STUDENTS AND THEIR SCORES ON THE COX-ORLEANS PROGNOSIS TEST OF TEACHING ABILITY

Cox-Orleans Prognosis	Student Teaching			Total f
	A	B	C	
375-381		1		1
368-374				
361-367				
354-360				
347-353				
340-346		3		3
333-339	2	1		3
327-332	3	2		5
320-326	1	11	1	13
313-319	4	2	2	8
306-312	3	1	1	5
299-305		2	3	5
292-298	1	4	1	6
285-291	2	3	1	6
278-284		2	1	3
271-277	3	2	1	6
264-270	1	4	1	6
257-263	2	3	1	6
250-256			2	2
243-249				
236-242				
229-235		1		1
Totals	22	42	15	79
M(Cox-Orleans)	303	304	290	301
r=.13 ± .074				

TABLE V

THE STUDENT TEACHING GRADES OF FORTY-  
SEVEN STUDENTS AND FACULTY ESTIMATES  
OF THEIR PROBABLE SUCCESS IN THE TEACH-  
ING PROFESSION

Faculty Estimates	Student Teaching			Total <i>f</i>
	A	B	C	
96-100		1		1
91-95	3	1		4
86-90	4	1		5
81-85	1	1	3	5
76-80	1	7		8
71-75	1	1	2	4
66-70		3	3	6
61-65		4	1	5
56-60		2	3	5
51-55	1		1	2
46-50			1	1
41-45				
36-40			1	1
31-35				
26-30				
21-25				
16-20				
11-15				
6-10				
1-5				
Totals	11	21	15	47
M(Estimates)	84	74	66	
r=.21±.09				

TABLE VI

THE STUDENT TEACHING GRADES OF NINETY-  
SEVEN STUDENTS AND THEIR GRADE POINTS  
PER HOUR OF CREDIT

Grade Points	Student Teaching			Total f
	A	B	C	
2.70-2.79	2			2
2.60-2.69	1	3		4
2.50-2.59	2	2	1	5
2.40-2.49	4	5		9
2.30-2.39		3		3
2.20-2.29	2	1		3
2.10-2.19	1	4		5
2.00-2.09		3		3
1.90-1.99	6	3	1	10
1.80-1.89	1	6	1	8
1.70-1.79		5	1	6
1.60-1.69	1	7	2	10
1.50-1.59		1	2	3
1.40-1.49	1	4	3	8
1.30-1.39	1	2	3	6
1.20-1.29	1	2	2	5
1.10-1.19		2	1	3
1.00-1.09		2	1	3
.90-.99			1	1
<b>Totals</b>		<b>23</b>	<b>55</b>	<b>19</b>
M(Grade Points)		2.13	1.88	1.50
r=.46 ± .05				1.87

TABLE VII

THE STUDENT TEACHING GRADES OF FORTY-TWO STUDENTS AND THEIR SCORES ON THE AMERICAN COUNCIL CIVICS AND GOVERNMENT TEST

Civics and Government Test	Student Teaching			Total f
	A	B	C	
170-179	1			1
160-169	1	1		2
150-159	2	1		3
140-149		3		3
130-139	3	3		6
120-129	1	3		4
110-119	2	4		6
100-109	1	1		2
90- 99	2	2		4
80- 89	1	1		2
70- 79	1	5	1	7
60- 69		1		1
50- 59				
40- 49	1			1
				
Totals M(Civ. & Gov.)	16 119	25 113	1 75	42 114
r = 11 ± .10				

TABLE VIII

THE STUDENT TEACHING GRADES OF THIRTY-EIGHT STUDENTS AND THEIR SCORES ON THE COLUMBIA RESEARCH BUREAU AMERICAN HISTORY TEST

American History	Student Teaching			Total f
	A	B	C	
130-136	1	1		2
123-129	1	2		3
116-122	1	1		2
109-115	1	2		3
102-108				
95-101	2	2		4
88- 94		4	1	5
81- 87	2	2	1	5
74- 80	1	3		4
67- 73	1	4		5
60- 66		2		2
53- 59	1	1		2
46- 52				
39- 45				
32- 38		1		1
 				
Totals M(Am. History)	11 97	25 87	2 88	38 90
r=.14 ± .11				

TABLE IX

THE STUDENT TEACHING GRADES OF EIGHTY-TWO STUDENTS AND THEIR SCORES ON THE COLUMBIA RESEARCH BUREAU ENGLISH TEST

Columbia English Test	Student Teaching			Total f
	A	B	C	
250-259		1		1
240-249	1			1
230-239				
220-229	2			2
210-219	1	4	1	6
200-209	4	3		7
190-199	1	4	2	7
180-189	2	5		7
170-179	1	4	1	6
160-169	4	4	3	11
150-159	1	6	2	9
140-149	1	5	2	8
130-139	1	4	2	7
120-129	1	2	2	5
110-119		2	1	3
100-109				
90-99	1			1
80-89			1	1
				
Totals	21	44	17	82
M(Columbia Eng.)	180	170	153	169
r = 26 $\pm$ .06				

relationship between grades in student teaching and the generalized measure of scholastic achievement represented by grade points. All three criteria, scatter of frequencies, means and the coefficient of correlation indicate such a relationship. The coefficient of correlation is the highest one found in this investigation. It is statistically significant and it has a forecasting efficiency of twelve per cent. These data would seem to support the policy that requires student teachers to have a specified average grade before they may become eligible for student teaching.

The other criteria of scholastic achievement, namely, scores on tests in American History, Civics and Government and English, yield correlations that are positive but so low that the tests have very little value in the prediction of grades in student teaching.

Basic Skills. The highest relationships found in this field were with reading vocabulary and reading comprehension. These are shown in Tables X and XII. Both coefficients of correlation are positive and high enough to meet the statistical criterion of significance. In fact, they compare favorably with the relationships found with scores on intelligence tests. The forecasting efficiency of the vocabulary test is three and that of comprehension is seven per cent.

The relationship between grades in student teaching and scores on the tests in language, arithmetic and work-study skills, are shown in Tables XII, XIII and XIV. The relationships are positive but too low to be of value in making predictions.

Personality Traits. The Allport A-S Reaction Study, Bell Adjustment Inventory, and Bernreuter Personality Inventory were used in studying relationships in this field. The results obtained are shown in Tables XV-XXIII. None of the correlations are high enough to have predictive value. The highest one is .17 and it exists between grades in student teaching and scores on the Dominance-Submission section of the Bernreuter Personality Inventory. This finding means, if meaning may be given a relationship so low, that there is a slight tendency for high grades in student teaching to be accompanied by dominance rather than submission in face to face situations.

The coefficients of correlation found in this investigation are summarized in Table XXIV. Only the highest four are statistically significant. Each of the remaining nineteen has a probable error that is more than one-fourth its coefficient of correlation.

Conclusions. Notwithstanding the limitations of this investigation, its findings support the following conclusions:

1. In general there is a low relationship between the grades of students in student teaching and measurements in each of the

TABLE X

THE STUDENT TEACHING GRADES OF EIGHTY-EIGHT STUDENTS AND THEIR SCORES ON THE IOWA EVERY PUPIL TESTS OF BASIC SKILLS,  
READING VOCABULARY

Iowa Reading (Vocabulary)	Student Teaching			Total f
	A	B	C	
50-51	2	4	1	7
48-49	12	15	4	31
46-47	7	16	4	27
44-45	1	5	1	7
42-43		5	1	6
40-41	1		4	5
38-39			1	1
36-37	1		1	2
34-35		1		1
32-33				
30-31				
28-29				
26-27				
24-25				
22-23		1		1
				
Totals	24	47	17	88
M(Iowa Reading)	48	47	45	47
r = 22 ± .07				

TABLE XI

THE STUDENT TEACHING GRADES OF EIGHTY-EIGHT STUDENTS AND THEIR SCORES ON THE IOWA EVERY PUPIL TESTS OF BASIC SKILLS,  
READING COMPREHENSION

Iowa Reading (Comprehension)	Student Teaching			Total <i>f</i>
	A	B	C	
80-82		2		2
77-79	1	4		5
74-76	1	5	1	7
71-73	3	3	2	8
68-70	5	6		11
65-67	6	7	1	14
62-64	2	5	4	11
59-61	3	1	2	6
56-58	2	2	2	6
53-55		3		3
50-52		3	2	5
47-49	1	1	1	3
44-46		1	1	2
41-43			1	1
38-40				
35-37		1		1
32-34				
29-31		1		1
26-28				
23-25	1	1		2
  				
Totals M(Reading) <i>r</i> =.33 ± .06	25 66	46 64	17 60	88 63

TABLE XII

THE STUDENT TEACHING GRADES OF THIRTY-  
SEVEN STUDENTS AND THEIR SCORES ON THE  
IOWA EVERY PUPIL TESTS OF BASIC SKILLS,  
IN LANGUAGE

Iowa Language	Student Teaching			Total f
	A	B	C	
284-286		2		2
281-283	1	1		1
278-280	2	2		4
275-277		4		4
272-274		2		2
269-271		3		3
266-268		1		1
263-265	1	1		2
260-262		3		3
257-259	2	3		5
254-256	1			1
251-253	1	1		2
248-250				
245-247	1			1
242-244	1			1
239-241		2	1	3
236-238				
233-235				
230-232		1		1
Totals	10	26	1	37
M(Iowa Lang.)	262	267	241	265
r=11±.11				

TABLE XIII

THE STUDENT TEACHING GRADES OF THIRTY-EIGHT STUDENTS AND THEIR SCORES ON THE IOWA EVERY PUPIL TESTS OF BASIC SKILLS IN ARITHMETIC

Iowa Arithmetic	Student Teaching			Total f
	A	B	C	
95-97	1	4		5
92-94	2	1		3
89-91		2		2
86-88		2		2
83-85		3		3
80-82	1	1		2
77-79		1		1
74-76		1		1
71-73	1			1
68-70		1		1
65-67	2	3		5
62-64	1	3		4
59-61		2		2
56-58	1		1	2
53-55	1			1
50-52		1		1
47-49				
44-46				
41-43	1			1
38-40		1		1
  				
Totals M(Arithmetic) r .013	11 72	26 77	1 58	38 75

TABLE XIV

THE STUDENT TEACHING GRADES OF THIRTY-EIGHT STUDENTS AND THEIR SCORES ON THE IOWA EVERY PUPIL TESTS OF BASIC SKILLS IN WORK-STUDY

Iowa Work-Study	Student Teaching			Total f
	A	B	C	
116-118		1		1
113-115		1		1
110-112		1		1
107-109		1		1
104-106	1	5		6
101-103	3	2		5
98-100		2		2
95-97	2			2
92-94	2	2	1	5
89-91		1		1
86-88		1	1	2
83-85	1	1		2
80-82			1	1
77-79	1			1
74-76	1	1		2
71-73				
68-70				
65-67				
62-64				
59-61				
56-58		1		1
53-55		1		
Totals	11	21	3	35
M(Work-Study)	94	96	88	95
r ,018				

TABLE XV

THE STUDENT TEACHING GRADES OF FORTY-FOUR STUDENTS AND THEIR SCORES ON ALL-PORT'S A-S REACTION STUDY (ASCENDANCE-SUBMISSION)

A-S Reaction Study	Student Teaching			Total f
	A	B	C	
48- 52		1		1
43- 47				
38- 42				
33- 37				
28- 32	1	3		4
23- 27	3	3		6
18- 22	2	1		3
13- 17	1	1		2
8- 12		3	1	4
3- 7		5		5
-2- -2	1	1		2
-7- -3	2	1	1	4
-12- -8		1		1
-17- -13	2	2	1	5
-22- -18				
-27- -23	1	2		3
-32- -28		1		1
-37- -33		1		1
-42- -38	1	1		2
Totals M(A-S Reaction) r .035	14 4	27 4	3 -3	44 4

TABLE XVI

THE STUDENT TEACHING GRADES OF FIFTY-  
 THREE STUDENTS AND THEIR SCORES ON "A",  
 HOME ADJUSTMENT, OF THE BELL ADJUST-  
 MENT INVENTORY

"A" Home Adjustment	Student Teaching			Total f
	A	B	C	
19	1			1
18				
17				
16		1		1
15				
14				
13				
12				
11				
10				
9	1	3		4
8	1	1		3
7	1	1		2
6		3		3
5	1	3		5
4	1	2		3
3	1	8		11
2	2	3		5
1	3	1		4
0	3	7		0
Totals	15	34	4	53
M(Home Adjust.)	4	4	5	4
r = .026				

TABLE XVII

THE STUDENT TEACHING GRADES OF FIFTY-THREE STUDENTS AND THEIR SCORES ON "B", HEALTH ADJUSTMENT, OF THE BELL ADJUSTMENT INVENTORY

"B" Health Adjustment	Student Teaching			Total <i>f</i>
	A	B	C	
16	1	1		2
15		1		1
14		1		1
13	1	1		2
12	1			1
11			1	1
10	1	2	1	4
9	1			1
8	1			1
7	1	3	1	5
6	1	2		3
5	1	5		6
4	2	2	1	5
3		7		7
2	2	8		10
1	2	1		2
0				1
<i>Totals</i>		15	34	53
<i>M</i> (Health Adjust.)		7	5	6
<i>r</i> .013				

TABLE XVIII

THE STUDENT TEACHING GRADES OF FIFTY-THREE STUDENTS AND THEIR SCORES ON "C", SOCIAL ADJUSTMENT, OF THE BELL ADJUSTMENT INVENTORY

"C" Social Adjustment	Student Teaching			Total f
	A	B	C	
30-31	1			1
28-29		1		1
26-27				
24-25	1	2		3
22-23				
20-21	1	1		2
18-19		1		1
16-17		1		1
14-15		2	1	3
12-13	3	3		6
10-11	1	3	1	5
8-9	1	1	1	3
6-7	2	7	1	10
4-5	1	4		5
2-3	2	7		9
0-1	2	1		3
Totals	15	34	4	53
M(Social Adjust)	11	10	11	10
r .035				

TABLE XIX

THE STUDENT TEACHING GRADES OF FIFTY-THREE STUDENTS AND THEIR SCORES ON "D", EMOTIONAL ADJUSTMENT, OF THE BELL ADJUSTMENT INVENTORY

"D" Emotional Adjustment	Student Teaching			Total f
	A	B	C	
20		1		1
19	1			1
18	1	2		3
17	2			2
16	1	2		3
15		1		1
14		2		2
13	1	2		3
12		1	1	2
11		1		2
10				
9				
8		4		
7	1			1
6	3	2		5
5		4		4
4	2	3		5
3	1	1		5
2	1	5		2
1	1	3	1	6
				41
Totals	15	34	4	53
M(Emotional Adj)	9	8	9	8
r .083				

TABLE XX

THE STUDENT TEACHING GRADES OF NINETY-TWO STUDENTS AND THEIR SCORES ON BI-N,  
NEUROTIC SYMPTOMS, OF BERNREUTER'S  
PERSONALITY INVENTORY

BI-N Neurotic Tendency	Student Teaching			Total f
	A	B	C	
200- 219				
180- 199				
160- 179			1	1
140- 159				
120- 139				
100- 119		1		1
80- 99				
60- 79				
40- 59	1	1		2
20- 39			1	1
0- 19		2	1	3
- 20- - 1		4		4
- 40- - 21	3	7	2	12
- 60- - 41		6	3	9
- 80- - 61	3	2		5
-100- - 81	4	5	2	11
-120- -101	1	4		5
-140- -121	1	5	3	9
-160- -141	4	7	2	13
-180- -161	2	7		9
-200- -181	2	2		4
-220- -201		1	2	3
Totals	21	54	17	92
M (BI-N)	-103	-90	-76	-91
r	.073			

TABLE XXI

THE STUDENT TEACHING GRADES OF NINETY-TWO STUDENTS AND THEIR SCORES ON B2-S, SELF-SUFFICIENCY, OF THE BERNREUTER PERSONALITY INVENTORY

B2-S Self-Sufficiency	Student Teaching			Total f
	A	B	C	
128- 142	1			1
113- 127		1		1
98- 112	1	4	3	8
83- 97	3	3	1	7
68- 82	1	3		4
53- 67	5	2	2	9
38- 52		7	3	10
23- 37	3	12	2	17
8- 22		5		5
- 7- 7	4	3	3	10
- 22- - 8		6	1	7
- 37- - 23		4	1	5
- 52- - 38	1	1		2
- 67- - 53	1	2	1	4
- 82- - 68		1		1
- 97- - 83				
-112- - 98				
-127- -113	1			1
Totals M (B2-S) r -.014	21 36	54 27	17 37	92 31

TABLE XXII

THE STUDENT TEACHING GRADES OF NINETY-  
 ONE STUDENTS AND THEIR SCORES ON THE  
 BERNREUTER PERSONALITY INVENTORY,  
 B4-D, DOMINANCE-SUBMISSION

B4-D Dominance- Submission.	Student Teaching			Total f
	A	B	C	
143- 157	1	5	1	7
128- 142	3	4	1	8
113- 127		5		5
98- 112	2	4	1	7
83- 97	2	5	4	11
68- 82	6	9	2	17
53- 67	2	6	2	10
38- 52		3	1	4
23- 37	2	1	2	5
8- 22	1	3		4
- 7- 7	1	4		5
- 22- - 8		3	2	5
- 37- - 23				
- 52- - 38		1		1
- 67- - 53				
- 82- - 68	1		1	2
<hr/>				
Totals	21	53	17	91
M (B4-D)	73	74	61	71
r = .168 ± .069				

TABLE XXIII

THE STUDENT TEACHING GRADES OF NINETY-TWO STUDENTS AND THEIR SCORES ON F2-S,  
SOCIAL ADJUSTMENT OF THE BERNREUTER  
PERSONALITY INVENTORY

F2-S Sociability	Student Teaching			Total f
	A	B	C	
98- 112		1	1	2
83- 97		1	2	3
68- 82	3	1		4
53- 67	1	2	1	4
38- 52	1	3		4
23- 37	2	6		8
8- 22	2	2		4
- 7- 7		5	2	7
- 22- - 8		4	5	9
- 37- - 23	2	3	1	6
- 52- - 38	6	5	1	12
- 67- - 53	2	7	2	11
- 82- - 68	1	5	1	7
- 97- - 83	1	5	1	7
-112- - 98		1		1
-127- -113	1	1		2
-142- -128				
-157- -143		1		1
  				
Totals	22	53	17	92
M (F2-S)	-19	-28	-10	-22
r .044				

TABLE XXIV

THE COEFFICIENTS OF CORRELATION BETWEEN GRADES IN STUDENT TEACHING AND TWENTY-THREE CRITERIA IN THE FIELDS OF INTELLIGENCE, TEACHING APTITUDES, SCHOLASTIC ACHIEVEMENT, PROFICIENCY IN BASIC SKILLS, AND PERSONALITY TRAITS

Criteria	r	P. E.
Grade Points. . . . .	.46	.05
Reading Comprehension . . . . .	.33	.06
Detroit Advanced Intelligence . . .	.28	.07
Am. Council English Test. . . . .	.26	.06
Henmon-Nelson Mental Ability. . . .	.23	.06
Reading, Vocabulary . . . . .	.22	.07
Faculty Estimates . . . . .	.21	.09
Bernreuter, Dominance-Submission	.17	.07
Am. Council Am. History Test. . . .	.14	.11
Cox-Orleans Teaching Prognosis. . .	.13	.07
O. S. U. Psychological Exam . . . .	.12	.06
Am. Council Civics and Govt . . . .	.11	.10
Iowa Every-Pupil Test. Language . .	.11	.11
Bell, Emotional Stability . . . . .	.09	
Bernreuter, Neurotic Tendency . . . .	.07	
Allport A-S Reaction Study. . . . .	.04	
Bell, Social Adjustment . . . . .	.04	
Bernreuter, Sociability . . . . .	.04	
Bell, Home Adjustment . . . . .	-.03	
Iowa Every-Pupil, Work-Study. . . .	.02	
" " Arithmetic. . . . .	.01	
Bernreuter, Self-Sufficiency. . . . .	.01	
Bell, Health Adjustment . . . . .	.01	

twenty-three measures employed in this study. A correlation of .46 is the highest one found.

2. The forecasting efficiency of each of the measures in predicting grades in student teaching is low, and, if the original assumption were correct, almost without value in foretelling success or failure in the teaching profession. Grade points in college work ranks highest and its efficiency is only twelve per cent better than pure guessing.

# A RESUME OF CURRENT TEACHER PERSONNEL RESEARCH

GLEN FULKERSON

Teacher Characteristics Study\*

University of California

Los Angeles, California

THE COMBINED factors of (a) the great increase in the pupil population without a comparable increase in the teacher population, and (b) the trend in American education toward higher training standards and more discriminating selection for every level and type of teaching personnel, have stimulated a growing interest in research dealing with teacher competencies, characteristics, and effectiveness. This interest has been noted particularly since the close of World War II.

By 1940 considerable attention had been accorded the delineation of desirable teacher traits and a certain amount of empirical study had been undertaken. By then, the carefully planned program of research and service of the American Council on Education's National Committee on Teacher Examinations was in full swing. This project, designed to assist in one phase of the teacher selection process — that of providing comparable estimates of teachers' intelligence, cultural background, awareness of current social problems, professional educational information, and mastery of academic subjects — now has been in operation fifteen years. Ryans, in various reports, of which the one published under the title "Appraising Teacher Personnel" (18) is typical, has summarized the activities of the program and analyzed certain of the Teacher Examination results.

In 1940, Barr (1) published summaries of 138 investigations that had sought, in various ways, to predict "teaching efficiency." In re-viewing the literature, he noted three basic approaches that had been employed in the attempted measurement of teaching ability: (1) the measurement of pupil change (only nineteen studies had tried to use this as a criterion); (2) the "static measurement-of-teachers' qualities" approach; and (3) the "dynamic-evaluation of performance" approach. Reporting again in 1952, Barr called attention to what he felt to be a shift in emphasis toward the last named approach with marked concentration of research effort upon the study of teacher-pupil relations and teacher personality. He also pointed out the long-neglected serious concern of researchers with the definition and identification of criteria of teaching efficiency. Domas and Tiedeman (8) in their annotated bibli-

\*The Teacher Characteristics Study, is sponsored by the American Council on Education and is conducted with funds provided by a subvention from the Grant Foundation.

graphy on teacher competence published in 1950, listed over a thousand titles, some 300 of which appeared between 1940 and 1950.

Research relating to teacher characteristics and teacher competencies probably is receiving more attention at the present than ever before. And the investigations being conducted appear, properly, to be more searching, more carefully designed, and more extensive than those undertaken two or three decades ago. Manifestations may be noted in several large-scale projects such as the Teacher Characteristics Study and the (New York) Cooperative Study to Predict Effectiveness in Secondary School Training, in programs of faculty and graduate research at such universities as Illinois, Minnesota, and Wisconsin, and in the concern shown by the American Educational Research Association in appointing its Committee on the Criteria of Teacher Effectiveness.

Despite the various early and more recent attempts to analyze teaching competency with a view to improving the preparation and selection of teachers, the results, in general, have been disappointing. There still seems to be no wholly satisfactory working plan or system that can be used by those who must make judgments about teacher effectiveness. This probably is due in part to the lack of researchers who are familiar enough with techniques of sampling, control, and statistical analysis to design studies that will yield usable results. Increased attention seems to be directed toward this problem, however, and techniques are becoming available and being utilized which it seems reasonable to expect may result in procedures of improved reliability and validity.

Together with the unavailability of research techniques, in contributing to the relative lack of significant conclusions about the identification and prediction of teaching competency, there has been the difficulty experienced by educators and educational researchers in trying to describe the essential nature of teaching effectiveness. It is significant that the American Educational Research Association saw fit to appoint the committee previously referred to, to study this problem. The first report of the AERA Committee on the Criteria of Teacher Effectiveness (16) was published in June, 1952 in the Review of Educational Research. It was hoped that this would serve to stimulate and guide educational research in this area.

#### An Effort on the Part of the Teacher Characteristics Study to Keep Abreast of Current Research

One of the most extensive of recent research programs devoted to the investigation of the criteria and dimensions of teacher behavior is the Teacher Characteristics Study. The Study is a six-year project under the auspices of the American Council on Education. It is sub-

sidized by The Grant Foundation. Although the Study began in 1948, the work has been conducted generally within a conceptual frame-of-reference similar to that suggested by the AERA Committee. The Teacher Characteristics Study research has been guided by two principal objectives, (1) the identification and analysis of patterns of teacher behavior, and (2) the development of materials useful for the prediction of teacher behavior. A number of related studies have been undertaken in light of these goals. Briefly they fall into two groups, (a) those of a descriptive nature which have to do with the identification of dimensions of teacher behavior, with conditions associated with teacher behavior, and with various comparisons of homogeneous teacher groups, and (b) those that are methodological, involving techniques for obtaining data, techniques for the analysis of data, and the development of various kinds of predictor material. Reports have appeared in the development of various kinds of predictor material. Reports have appeared in the literature from time to time describing both the general progress of the Study and the results of some of the completed phases. (20) (21) (22) (23) (24) (25)

It is a direct interest, as well as an obligation, of any alert and forward looking research program to maintain familiarity with related research being conducted. Unfortunately, the publication lag in research journals makes it difficult to maintain close contact with new findings and techniques. In an effort to keep abreast of investigations that are currently in progress (and which may not appear in print for several years due to delayed publication) the Teacher Characteristics Study staff undertook, and completed in the early summer of 1953, a nation-wide survey of research in progress or research only recently completed pertaining to teacher characteristics, teacher competencies, and teacher effectiveness. About 600 members of the American Educational Research Association were circularized in view of their presumed interest in the problem and the likelihood that they would have knowledge of persons who were conducting research in the area. They were asked to advise the Teacher Characteristics Study staff of public school personnel, college or university faculty, or graduate students who might be engaged in such investigations. As a result of the inquiry, the names of some ninety individuals were obtained. These persons were subsequently requested to provide information about their research as to: purposes; hypotheses; variables studied; control techniques; characteristics of population and sample studied; raters of observers employed; nature of and reliability and validity of scales or instruments used; statistical techniques employed in analyzing data; major findings; and possible implications of the results.

#### Summary of Research in Progress Dealing with Teacher Characteristics and Behavior

A review of the results of the Teacher Characteristics Study's

inquiry into research in progress, as revealed by the research summary sheets returned seems to support the observation that the current main emphasis is on aspects of teacher personality and teacher behavior, although study of other phases of the teacher personnel problem also is being pursued.

The work that is presently underway appears to be spearheaded by a limited number of carefully designed investigations, or groups of coordinated investigations, that give promise of significant results. But these are in the minority. A larger number of studies reported might be classified as limited in scope, with less attention given to control factors, defined criteria, and other necessary considerations.

What seems to be one of the major efforts to contribute to a better understanding of the overall problem of teaching effectiveness is the Cooperative Study to Predict Effectiveness in Secondary School Teaching, directed by Paul A. Hedlund for the New York State Education Department. (11) Still in progress, this study was undertaken at the request of twenty-four upstate New York Colleges and universities preparing secondary school teachers in an effort to gain information that would aid them in selecting students for teacher training programs. Eighteen of the colleges and universities have cooperated actively in the study from its beginning in 1948. This study attempts to answer the principal question: What student characteristics, measurable at the time of admission to a program of professional education, are related to success in that program and to effectiveness in teaching? The answer to this question is being sought by: (1) gathering a comprehensive body of information on all applicants for teacher education in the cooperating colleges; (2) following the progress of the students through college and into employment to note success in the teacher education program and effectiveness during the first year of teaching; and (3) analyzing the prediction data in relation to academic and teaching success to determine the factors which colleges should consider in selecting their education students.

The basic prediction data include extensive biographical information; information regarding personal and social characteristics of the entering student as revealed by standardized inventories, interviews, and ratings by former teachers, present teachers, and speech experts; and measures of aptitudes and interests obtained from grade transcripts and a battery of psychometric tests and inventories. The third-year progress report of the Cooperative Study indicates a number of factors which tentatively seem to have predictive value. A Prediction Index, which will be refined further during the remainder of the study, has been developed on the basis of the findings to date. The study is confining its research to secondary school teachers.

At the elementary school level a pilot project is being conducted by the Bank Street College of Education in New York under the direction of Barbara Biber. (2) The investigation is sponsored by the Sub-

committee on Teacher Selection of the Citizens' Committee on Children of New York City in cooperation with the Board of Examiners of the Board of Education of New York City. In broad outline the study involves the development of a set of instruments designed to measure personality characteristics related to responsibilities of teachers in the elementary grades, measurement and follow-up by observation and interview of those teachers who receive appointments, and analysis of the relationships between the two sets of data in an effort to determine which approaches to assessment yield effective indices of selection. The test battery employed includes a cartoon test, an authoritarian scale, an incomplete sentence test, a draw-a-teacher technique, and an autobiographical questionnaire. These tests were administered to some 1600 applicants for the New York city schools in February, 1953 as part of the regular examination procedure. The follow-up sample will be approximately 200 teachers who were recently appointed to teaching positions.

Travers, Rabinowitz, Page, Neomvicher, and Ventur (26) reported in March, 1953, the results of seven different exploratory studies in teacher personality conducted for the Division of Teacher Education at The College of the City of New York. Two of these dealt with performance on the Rorschach ink blot test as related to student teaching and the effects of mental set on certain aspects of Rorschach performance. Three other studies investigated a picture drawing technique for studying certain outcomes of teacher education and the use of drawings as a screening device. One study reported on the anxieties of a group of student teachers, and still another discussed the Q-Technique in analyzing personality and its application to teacher education.

Another exploratory study conducted for the purpose of better selecting and predicting the effectiveness of student teachers has been carried on for the last three years at the University of Illinois by Merle M. Ohlsen and Raymond E. Schultz. (15) Team judgments of supervisors were used as a criterion for obtaining groups of best and poorest student teachers. These groups are being compared on the basis of: Alexander's adaptation of the Thematic Values Test, a Sociometric Test, the Strong Vocational Interest Blank, Eberman's Trait Universe, sentence completion tests, grade point averages, and supervisors' analysis of strongest and weakest points in student teachers. Five papers giving the results have been accepted for publication.

Paul R. Grim, University of Minnesota, is engaged in the construction of appraisal instruments for evaluating growth in teaching competency as revealed through pupils' ratings and principals' ratings. (10) Two psychometric instruments have been used, the Student Reaction Inventory and the Teacher Characteristic List. Data to validate the two instruments were gathered from ten high schools located in various parts of the U.S. In each school the principal was asked to select ten teachers, the five considered "weakest" and the five considered

"strongest." The SRI was administered to the classes of the teachers, and the principal rated the teachers on the TCL. The author anticipates, that with further refinement, the scales may contribute as one of a battery of instruments used to evaluate teaching.

The components of "causality of approach" (not specifically defined in the research summary submitted) in elementary school teachers is the subject of an investigation being carried out by Eugene E. Levitt of the Child Welfare Station, State University of Iowa. (12) Levitt hypothesizes that "causality of approach" — which is to be described in detail in a forthcoming publication — is related to certain personality characteristics among which are emotional stability, permissiveness, tolerance of ambiguity, and flexibility. Twenty-seven fifth and sixth grade teachers are being rated on ten hours of classroom observation by each of three observers on a six-point scale of "causality." It is anticipated that a number of formal tests will be used and the correlation between the test results and the observations will be analyzed with the expectation of devising a battery of easily administered scales which will successfully discriminate "causality" in elementary school teachers.

The relationship between the personal characteristics of teachers, as measured by the Kuder Preference Record, and teaching success, as measured by superintendents' ratings, was ascertained by Sandford S. Davis in a doctoral study at Iowa State Teachers College. (7) The sample is composed of Iowa elementary teachers who have completed the two-year elementary training program at Iowa State Teachers College and have completed one year of training.

Walter Cook has continued investigations of the Minnesota Teacher Attitude Inventory. Normative data, obtained from a ten percent random sample of all the teachers in Minnesota, provide the basic data for analyses now in progress that seek relationships between teacher attitudes and items appearing in the Minnesota Multiphasic Personality Inventory. Low and high groups from this sample were given the MMPI and an attempt is being made to determine to what extent the MMPI can be used in predicting teacher-pupil rapport. (6) Thus far, some 250 items on the MMPI have been identified which discriminate between the low and high teachers. Two or three new scales using these items are being developed.

The research office of the American Association of Junior Colleges, is currently studying two groups of junior college teachers to see if significant differences may be found with respect to certain aspects of preservice preparation, "life styles," community-life participation, and methods of teaching. (5) C. C. Colvert and M. L. Litton are conducting the study. The sample is made up of instructors nominated from 255 junior colleges in the U.S. Each college nominated a "good" teacher. Another teacher was selected at random from the faculty of each of the 255 colleges to be a member of the "control"

group. Questionnaires and a transcript of college courses were obtained from each member of the two groups and the data analyzed. Only the first part of the study, that of preservice preparation has been completed at this time. The evidence thus far does not appear to substantiate the previously held assumptions and opinions of administrators, teachers, and specialists as to the desirable preservice preparation of junior college teachers.

Norman D. Bowers at the University of Minnesota is concerned with the construction of an inventory that will appraise certain aspects of job satisfaction of teachers with a view toward helping eliminate the exodus from the teaching profession. (3) After the items for this inventory have been obtained, they will be administered to a group of teachers who have been selected on the basis of satisfaction with teaching and to another group of teachers who have left the profession within the past two years because of stated dissatisfaction with the profession. Item analysis techniques will be utilized and a scoring key developed.

The development of a standard interview technique for the selection of teachers, using "Kodaslides" as stimuli for inventory response is being studied at Boston University by Charles O. Richter. (17) Some 200 slides showing selected teaching and learning situations are to be item-analyzed and validated. Each teacher responds or reacts to the picture, indicating, among other things, the degree of his approval or disapproval of the practice illustrated. In addition to giving responses to the slides, each teacher fills out an appraisal sheet. This sheet will have five or six choices, one of which will be a high-level response or choice, and the others will be lower level responses. A group of known strong teachers and a group of known weak teachers will be used to determine the discriminating value of the slides and to classify the responses on the appraisal sheet into high and low levels.

A recent doctoral dissertation of Arnold Buchheimer (4) at Ohio State University dealt with behavior dimensions of College of Education freshmen. The research proposed to answer three questions: (1) What dimensions can be isolated by an analysis of test behavior used to assess students' potentialities as teachers? (2) What hypotheses for further study will such an analysis yield? (3) To what extent are the dimensions of behavior of students in an introductory survey course in education similar to the behavioral dimensions of successful teachers, as such dimensions have been suggested by other studies? In attempting to answer the first question, a factor analysis was carried out resulting in the abstraction of eight factors defined as: general scholastic potential and general ability; general adjustment; social awareness and social participation; verbal facility; mathematics and reasoning ability; "studies concerns"; social inadequacy; and "no problems." The relationships of behavioral dimensions of students to those of successful teachers are clouded by the complexity of the criteria of suc-

cessful teaching, and apparently depend on which criteria are utilized.

An investigation of the variability in observed classroom behaviors of junior high school teachers and classes has just been concluded by Edwin Wandt and Leonard M. Ostreicher (27) at the College of the City of New York. Repeated classroom observations of teachers and classes were made for the purpose of determining: (1) the consistency of classroom behaviors of teachers and classes over a period of time; (2) the relationship of the teachers behaviors to the characteristics of the particular class observed; and (3) the representativeness of initial observations of the average of observations made over a longer period of time.

The instrument employed consisted of fourteen rating scales designed to assess the social-emotional climate of the classroom based on the behaviors of the teachers and the class. Analysis of the data suggested that: (1) social-emotional climate in the classrooms of observed teachers may vary widely from occasion to occasion; (2) in some cases, at least, the climate may vary systematically with the type of class observed; and (3) initial observations may be very unreliable indices of the "typical" climate, even when the type of class is held constant.

Wilbur Williams (28) of Michigan State Normal College has underway a four-year study dealing with the evaluation and prediction of student teaching effectiveness and on-the-job teaching success. His intent is to discover traits and characteristics which will be useful in identifying those students who are apt to become outstandingly good and poor teachers. To accomplish this, Williams has taken a random sample of 100 elementary and 100 secondary teacher trainees. A battery of standardized tests was given these two groups. In addition, employing officials were asked to predict on the basis of interviews with the students their likely success in student teaching and on-the-job teaching. Supervising teachers also were asked to rate the student teachers for on-the-job teaching. Observers will be trained to follow up these subjects, who have completed one year of teaching to discover differences in teaching effectiveness.

#### Miscellaneous Peripheral Investigations

A number of studies are in progress that are not directly and solely concerned with the problem of teacher competency, but are tangentially related to the problem.

As one phase of the Kellogg Project (a regional study encompassing several states) O. B. Graff, at the University of Tennessee is in charge of an investigation which seeks to identify the characteristics of effective administrative leadership in education. (9) As an hypothesis for the planning of the project, the staff developed a statement of char-

acteristics which probably would be common to both good teaching and good school administration. A rating instrument also was developed, permitting the assessment of these characteristics on a five-point scale, which it was hoped would serve as a useful device in evaluating the data gathered about students in the educational administration program. It is anticipated that data from on-campus observations and on-the-job observations would be analyzed. Leaderless group discussions, free-response autobiographies, personal data forms, personal interviews, and problem report forms were other methods used in obtaining information about the characteristics of the students making up the population.

In cooperation with the Psychology department at the University of Tennessee, Graff also obtained data on prospective administrators resulting from administration of the Minnesota Teacher Attitude Inventory, the Allport-Vernon Study of Values, the Cooperative English Test, an opinion scale, an attitude questionnaire, the Kuder Preference Record (vocational), the Watson-Glaser Critical Thinking Appraisal Test, the Miller Analogies test, and the Rorschach test. The two sets of data will be appraised separately in relation to the criteria represented in the rating scale. Future plans of the project call for extensive and intensive follow-up field visits with the participants after they assume active administrative responsibilities to attempt to determine the validity of predictions made in the earlier study.

Chester McNerny of Pennsylvania State College (14) is attempting to determine which administrator and teacher characteristics seem to impede teacher growth and hinder the improvement of the educational process. It is expected that suggestions for improvement will grow out of the study.

In an effort to develop a better basis for establishing state and local certification standards and a better basis for developing teacher education programs, the Office of Education, Washington, D. C. is currently engaged in a study of competencies needed by teachers of exceptional children. (13) The study, which ~~is directed~~ by Romaine P. Mackie, has two aspects: (1) an analysis of status data; and (2) a survey of the opinion of experts. Committees of the latter have been set up for thirteen programs of education for exceptional children. Inquiry forms have been developed for each. The data are now being collated, and a final report is expected in 1954.

The Advisory Educational Group of the Metropolitan Life Insurance School Health Bureau proposed a study in 1951 of the methods as used to determine the physical fitness, emotional stability, and mental health of students selected for admission to the teacher-education program of schools in the United States and Canada. (29) A committee appointed to conduct the study consisted of W. Carson Ryans, University of North Carolina, J. Paul Leonard, San Francisco State Col-

lege, and Robert H. Roberts, London, Ontario, Canada. Questionnaires sent to 1,235 teacher education institutions resulted in a fifty-one percent response in the U. S. and a seventy percent response in Canada. The study is concerned with six aspects of fitness for teaching: academic interest and accomplishment; wholesome personality; mental attitudes; physical fitness; capacity for growth; and interest in all-round activities. A brief summary report is now available.

### Conclusion

Research dealing with teacher characteristics, competencies, and effectiveness over the last several years has suggested that teaching ability is probably a complex of abilities rather than a unitary trait or behavior pattern. A more fruitful approach has seemed to be the hypothesizing of teaching ability in multidimensional terms that take into account the possibility of different patterns of effective teacher behaviors for different kinds of teachers, different kinds of pupils, and different kinds of educational situations. Most investigators seem to feel that the various aspects of teaching ability should be studied independently before attempting to study such abilities in the composite. Such an orientation is evinced by some of the studies briefly summarized here. Several of them concern themselves with the development of psychometric instruments that will predict aspects of teacher effectiveness related to teacher personality.

The varied attempts of the research in progress will doubtless, when completed, contribute fragments of information that will help lead to a more complete understanding of teacher effectiveness, but it seems apparent that research which will provide more than a partial and tentative answer regarding effective teaching is still to be undertaken. Based on the brief information provided for the foregoing account, it appears that researchers are still plagued by the problems of controlling relevant variables, and the difficulty of establishing adequate validating criteria.

It is also noticeable that many of the studies are concerned with local problems somewhat limited in scope which may or may not have value for more general application. Since the determination of teacher characteristics, competencies, and effectiveness poses one of the most crucial problems facing educational researchers today, it is imperative that means and methods be found to conduct more extensive, comprehensive, and basic research.

## REFERENCES

1. Barr, A. S. "Measurement of Teaching Ability." Review of Educational Research, 10:182-85, June 1940.
2. Biber, Barbara "Pilot Study in Teacher Selection." (Now underway at Bank Street College, New York.)
3. Bowers, Norman D. The Development and Validation of an Inventory Designed to Appraise Certain Aspects of Job Satisfaction. (Now underway at University of Minnesota, Minneapolis, Minn.)
4. Buchheimer, Arnold Performance Dimensions of Education Freshmen. Doctoral dissertation at Ohio State University, Columbus, Ohio.
5. Colvert, C. C. and Litton, Maurice A Study of Two Groups of Junior College Teachers to See if There Are Significant Differences with Respect to Certain Aspects of Pre-Service Preparation, Life Styles, Community Life Participation, and Methods of Teaching. (Now underway at the University of Texas, Austin, Texas.)
6. Cook, Walter W. (Research now underway at University of Minnesota, Minneapolis, Minn.)
7. Davis, S. S. The Relationship Between School Superintendents' Ratings of Elementary Teachers and the Kuder Preference Record-Personal and other Measured and Rated Teacher Characteristics. (Doctoral Dissertation at University of Colorado, Boulder, Colorado.)
8. Domas, S. J. and Tiedeman, D. V. "Teacher Competency: An Annotated Bibliography." Journal of Experimental Education, 19, December 1950.
9. Graff, O. B. The Characteristics That Are Effective to Administrative Leadership in Education. (Now underway at the University of Tennessee, Knoxville, Tennessee.)
10. Grim, Paul R. Construction of Appraisal Instruments for Evaluating Growth in Teaching Competency. (Now underway at University of Minnesota, Minneapolis, Minn.)
11. Hedlund, Paul A. Cooperative Study to Predict Effectiveness in

Secondary School Teaching. Third Progress Report. The State Education Department, Albany, New York, 1953.

12. Levitt, Eugene E. The Components of "Causality of Approach" in Elementary School Teachers. (Now underway at Child Welfare Research Station, State University of Iowa, Iowa City, Iowa.)
13. Mackie, Romaine P. Qualifications and Preparation of Teachers of Exceptional Children. (Now underway at Office of Education, Washington, D. C.)
14. McNerney, Chester. What Types of Teachers and Administrators Are A Hindrance to the Improvement of the Educational Process, and What Can Be Done for Them. (Now underway at Pennsylvania State College, State College, Penna.)
15. Ohlsen, M. M. and Schultz, R. E. Characteristics of Best and Poorest Student Teachers. University of Illinois, Urbana, Illinois, 1953.
16. Remmers, H. H., (Chairman.) "Report of the Committee on the Criteria of Teacher Effectiveness." Review of Educational Research, 22: 238-63, June 1952.
17. Richter, C. O. A Study in the Use of Slides for Teacher Selection. (Now underway as a doctoral research at Boston University, Boston, Mass.)
18. Ryans, D. G. "Appraising Teacher Personnel." Journal of Experimental Education, 16:1-27, Sept. 1947.
19. Ryans, D. G. "The Use of the National Teacher Examinations in Colleges and Universities." Journal of Educational Research, May, 1950.
20. Ryans, D. G. "The Criteria of Teaching Effectiveness," Journal of Educational Research, 42: 690-699, May 1949.
21. Ryans, D. G. "A Statistical Analysis of Certain Educational Viewpoints Held by Teachers," Journal of Experimental Education, 22: 119-32, December, 1953.
22. Ryans, D. G. "A Study of the Extent of Association of Certain Professional and Personal Data with Judged Effectiveness of Teacher Behavior." Journal of Experimental Education, 20: 67-77, Sept. 1951.

23. Ryans, D. G. "A Study of Criterion Data (A Factor Analysis of Teacher Behaviors in the Elementary School.)" Educational and Psychological Measurement, 12 No. 3 :333-44, 1952.
24. Ryans, D. G. and Wandt, E. "Investigations of Personal and Social Characteristics of Teachers." Journal of Teacher Education, 3:228-231, Sept. 1952.
25. Ryans, D. G. "The Investigation of Teacher Characteristics." The Educational Record, 34:371-96, October, 1953.
26. Travers, Robert M.W., et al. Exploratory Studies in Teacher Personality. Publication 14, College of the City of New York, Division of Teacher Education Office of Research and Evaluation, March, 1953.
27. Wandt, Edwin and Ostreicher, Leonard M. Variability in Observed Classroom Behaviors of Junior High School Teachers and Classes. Publication 16, College of the City of New York, Division of Teacher Education Office of Research and Evaluation, June, 1953.
28. Williams, Wilbur A. "Evaluating and Predicting Student Teaching and On-the-job Teaching Success." (Now underway at Michigan State Normal College, Ypsilanti, Mich.)
29. Fitness for Teaching. School Health Bureau, Metropolitan Life Insurance Company, 24: Nos. 3 and 4, March and April 1953.

# EFFECTIVE COLLEGE TEACHING: A PRELIMINARY CAUSAL ANALYSIS

ROBERT M. FRUMKIN  
Ohio State University  
WILLIAM H. HOWELL  
South Carolina State College

## Introduction

THE IDEA for this study was stimulated by the current national interest in intellectual dishonesty on the part of the students in our American colleges. One of the writers believed that too little has been made of the intellectual lethargy characteristic of some professors in our colleges. This then led to an interest in what effective teaching in college is and what may be the causes which lead a full professor to be effective or ineffective in his teaching.

## Formulation of the Problem

1) Long range problem: to find the precise conjuncture and organization of factors to which the difference, i. e., the difference between an effective and ineffective teacher, belongs.

2) Immediate problem: to begin a series of research studies which will make the solution to the former problem more accessible. This study represents a first attempt toward the accomplishment of this goal.

## Major Hypotheses

I. That aside from scholarship, it is personality factors which make the difference between the effective and ineffective college professor.<sup>1</sup>

II. Causal Hypothesis: it is a conjuncture and organization of scholarship, personality, and other factors which make the difference between an effective and ineffective teacher.

1. This hypothesis has been suggested by H. Gordon Hullfish, Professor of Education, Ohio State University.

Methodology

1. Selection of Professors. Five graduate students in the physical sciences and five graduate students in the social sciences were asked to select five full professors they believed to be effective teachers and five full professors they believed to be ineffective as teachers. In this way there should have been, theoretically, 50 effective and 50 ineffective professors written on 100 slips of paper. After, however, eliminating duplicates and inconsistencies<sup>2</sup> (a professor chosen as effective and ineffective both) the field was narrowed down to 83, 48 alleged effective and 35 alleged ineffective. From each of the two separate piles of slips placed in separate boxes one slip was picked by a non-participating student. In this way we had selected our effective and ineffective full professors for this pilot study. As it turned out, our effective professor represented the physical sciences. From now on he will be referred to as Professor Y. The ineffective professor represented one of the social sciences. From now on he will be referred to as Professor X.<sup>3</sup>

2. Ratings of Graduate Students as to the Effectiveness of the Teaching of Professor Y and Professor X: Graduate students in several classes in the respective departments of Professor Y and Professor X were asked to rate the professors in question on the basis of the Missouri Criteria of Effective College Teaching.<sup>4</sup> The students were asked not to rate the professor unless they had had at least 4 quarter hours of work under him,<sup>5</sup> 3 quarter hours of which consisted of a lecture course. In this manner, and with the cooperation of our colleagues, we were able to get 27 ratings for Professor Y and 34 ratings for Professor X.<sup>5</sup>

3. Total Mean Score Rating of Effective Teaching for each Professor: The Missouri Criteria of Effective College Teaching consists

2. Students in our classes suggested that it would be interesting to investigate these inconsistencies. These inconsistencies suggest that a professor may teach one course very well and another very poorly. He may be better as a lecturer in advanced courses than as one in open courses, etc.

3. The titles, "Professor Y" and "Professor X", have been used in order not to reveal the identity of the two professors. For this reason the data given in the social histories of these two men were presented in a disguised but meaningful manner.

4. The Criteria of Effective College Teaching appearing in Figure 1 are the result of the work of the Committee on the Improvement of Instruction of the College of Arts and Science of the University of Missouri, appointed by Dean Ellis in 1947.

5. This entire pilot study was carried on at Ohio State University.

of 25 items (See Figure 1, The Rating Sheet for Effective Teaching), each of which is answered on an 8 point scale. By adding the ratings of each graduate student on each item for each professor it was possible to get a mean score rating of effectiveness on each item for each professor, and also a total mean score rating of teaching effectiveness by adding the mean score ratings on each of the 25 criteria and dividing by 25.

4. Discrepancy Scores as an Index of Significant Differences Between Professors Y and X in Teaching Effectiveness Criteria: The writers felt that a discrepancy score on each item would be a more valuable tool for finding out which criteria were most important in differentiating the effective from the ineffective professor. For this reason the mean score rating on each item (or criterion) for the effective professor was subtracted from the mean score rating on each criterion for the ineffective professor.

#### Social History of the Alleged Effective Professor Y<sup>6</sup>

- a. Age. He is in the middle fifties.
  - b. Birthplace. He was born and raised in the eastern part of the U. S. in a large urban community.
  - c. Grandparents. They were professors and schoolteachers.
  - d. Parents. His father was a professor; his mother was a schoolteacher.
  - e. Siblings. Is one of several children; one sibling is in business, the other is a school teacher.
  - f. Education. He received his Ph.D. in a large eastern U. S. University.
  - g. Work History. Has held government research positions, and is now the member of a very important U. N. Commission.
  - h. Scholarship. His contributions to the field in which he works are so important that they are required reading of all graduate students in that field. In the past few years he has also given many of the lectures in the distinguished lecture tours made in his field.
  - i. Reasons for going into teaching. His father who was a professor set a most favorable example. He feels that in teaching young people he can help them to become good thinkers and good citizens.
6. Both professors were approached in the following way: "You, Sir, have been chosen among many to be the object of a research project which has as its purpose the study of the social factors relevant in a man's rise to the rank of full professor. We would appreciate your cooperation in this project." Fortunately the two professors in question were very cooperative. The writers also obtained further information about these men from Who's

FIGURE I. THE RATING FOR EFFECTIVE TEACHING DIRECTIONS:

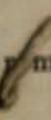
Directions: Please rate the professor in question on each of the following 25 qualities by circling the appropriate number at the right of that quality. 8 is the highest possible rating and 0 the lowest.

1. Preparation for class meetings (meetings planned with care, sometimes planned, etc.)	8-7-6-5-4-3-2-1-0
2. Teacher's interest in subject.	8-7-6-5-4-3-2-1-0
3. Reaction of students (students are alert and attentive, slightly, not at all.)	8-7-6-5-4-3-2-1-0
4. Classroom management (attends to the comfort of students as to light, temperature, etc.)	8-7-6-5-4-3-2-1-0
5. Scholarship (knowledge of subject seems broad and accurate, knowledge deficient, etc.)	8-7-6-5-4-3-2-1-0
6. Ability to express thought (mastery of language; meaning always clear, etc.)	8-7-6-5-4-3-2-1-0
7. Voice (well adapted to classroom diction, etc.)	8-7-6-5-4-3-2-1-0
8. Spirit of growth (presentation fresh and vital).	8-7-6-5-4-3-2-1-0
9. Assignments (given clearly and carefully, etc.)	8-7-6-5-4-3-2-1-0
10. Class discussion (discussions valuable, etc.)	8-7-6-5-4-3-2-1-0
11. Feeling between teacher and class (strong atmosphere of mutual goodwill, etc.)	8-7-6-5-4-3-2-1-0
12. Sense of humor (keen and appropriate sense of humor, etc.)	8-7-6-5-4-3-2-1-0
13. Self-confidence (sure of himself, poised, etc.)	8-7-6-5-4-3-2-1-0
14. Tolerance and mental flexibility (welcomes differences of opinion with respect, etc.)	8-7-6-5-4-3-2-1-0
15. Personal appearance (good grooming, taste, etc.)	8-7-6-5-4-3-2-1-0
16. Personal mannerisms (freedom from them, etc.)	8-7-6-5-4-3-2-1-0
17. Conduct of class (conducted in competent, businesslike manner; confusion, etc.)	8-7-6-5-4-3-2-1-0
18. Efficiency in procedure (time well-utilized)	8-7-6-5-4-3-2-1-0
19. Approachability (teacher welcomes conferences and tries to be understanding and helpful, etc.)	8-7-6-5-4-3-2-1-0
20. Adjustment to level of class (instruction followed by most of class; not always, etc.)	8-7-6-5-4-3-2-1-0

- |  |                   |
|--|-------------------|
| 21. Stimulus to thinking (presents an abundance of novel and stimulating ideas and viewpoints, etc.)   | 8-7-6-5-4-3-2-1-0 |
| 22. Validity of Examinations (exams correspond closely to course materials; not always, etc.)  | 8-7-6-5-4-3-2-1-0 |
| 23. Use of examination as a teaching device (points out strong points and weak points of exams, makes clear the basis of grading, helping students to succeed better thereafter, etc.)   | 8-7-6-5-4-3-2-1-0 |
| 24. Major objectives of course (objectives of the course are clearly explained and made obvious throughout course, etc.)   | 8-7-6-5-4-3-2-1-0 |
| 25. General Rating of teacher (if all the teachers with whom I have taken at least 4 quarter hours of work, including 3 quarter hours, of lecture work, all qualities considered, were divided into eight groups, I should place this one as indicated above.) | 8-7-6-5-4-3-2-1-0 |

---

This space is reserved for any additional comments you may care to add concerning the professor in question:

Note: Please do not place your name or any other identifying marks on this questionnaire. Thank you for your ~~concern~~ .

\*Evaluation of these comments will appear in a forthcoming paper.

- j. Methods of teaching effectively. Professor Y believes that in order for one to be an effective teacher one must know his subject thoroughly. One must love the subject he teaches and transmit this love to his students. An effective teacher must get his students to think; especially to think critically and clearly. An effective teacher must keep up with the times. One must try different ways of communicating effectively with one's students.

#### Social History of the Alleged Ineffective Professor X

- a. Age. Professor X is in his late fifties.
  - b. Birthplace. He was born in a midwestern city. He spent most of his life in this city.
  - c. Grandparents. His grandparents were engaged in business.
  - d. Parents. His parents were also engaged in business.
  - e. Siblings. He is one of several children. All of his siblings are in business; he is the only one in his whole family not in someway connected with the business world.
  - f. Education. He received his education in America and studied a few years in Europe after receiving his Ph. D. from a university on the east coast. Although his major field of concentration was in the social sciences he devoted a great deal of time to the study of literature.
  - g. Work History. He has taught in several universities. He has also engaged in government research work.
  - h. Scholarship. Although he has published much of his work his contributions to his field have not been of the same importance which the work of Professor Y has so clearly demonstrated in his field.
- Reason for going into teaching. He said he went into teaching because he wanted to be successful in a field that was not competitive. ~~He feels~~ that in teaching one's success is not determined by another person's loss or failure. He seems to hate anything to do with business.
- j. Methods of teaching effectively. He suggests that an effective teacher must encourage young to locate, use, and evaluate relevant materials as to the background of class instruction. Students must also be taught to identify and analyze significant

Who, Men of Science, etc. From the libraries of Ohio State University lists of publications by these professors were obtained. The titles of these works cannot be revealed because of our promise to keep the identity of these men unknown. However, let it be known to the reader that these works were evaluated by experts in these fields and taken into consideration in the writers' interpretation of the data.

TABLE I

## MEAN SCORE RATING ON EACH CRITERION OF EFFECTIVE TEACHING

Item No.	M. S. R. *	M. S. R.
	Prof. Y	Prof. X
1.	7.80	1.20
2.	7.65	3.10
3.	6.30	2.05
4.	6.75	2.10
5.	7.90	6.05
6.	7.25	4.40
7.	5.90	6.20
8.	6.95	1.05
9.	7.75	3.15
10.	5.90	2.30
11.	6.55	3.55
12.	6.30	6.05
13.	6.70	6.15
14.	7.70	6.25
15.	7.95	4.05
16.	6.35	6.15
17.	7.80	3.05
18.	7.90	5.25
19.	6.75	7.10
20.	6.90	7.05
21.	7.65	2.65
22.	7.75	1.50
23.	7.60	4.25
24.	7.80	4.05
25.	7.95	3.45

\*M. S. R. refers to mean score rating.

and contemporary problems and form judgments about them.

Professor Y had an average mean score rating of 7.19. Professor X had an average mean score rating of 4.09. The difference between the two means was 3.10. While an examination of this table would show some apparent differences it is felt that a more discerning picture may be derived from a table showing the rank of the criteria according to the discrepancy scores. Hence, Table II follows.

If our first hypothesis is valid, that is, that aside from scholarship, personality factors make the difference between an effective and an ineffective professor, Table II should reveal a greater discrepancy score between the Mean Y and Mean X for the criteria tends to indicate that the ten greatest discriminating items do not seem to be related to personality but the ten which show the least discrepancy seem to be more directly related to personality. Thus, from this examination of our data, our first hypothesis is not substantiated.

The second hypothesis, that is, that it is a conjecture and organization of both scholarship, personality, and other factors which make the difference between an effective and an ineffective professor can now be examined. While this hypothesis has been partially invalidated by the above remarks certain observations should be made before proceeding further. The verification of a hypothesis of causation is conducted on two levels. We apply statistical analysis to see how one phenomenon is associated with another, and we seek to discover whether the association of phenomena is meaningful. Since our statistics do not seem to aid in our interpretation, which may be due to limitations which will be mentioned later in this paper, we seek for an explanation from our experience as members of similar cultural complexes. We take the task of projecting ourselves by sympathetic reconstruction into the situation as it is assessed by others. Experience and knowledge provide us with backgrounds into which we seek to fit the specific phenomenon.

On the basis ~~of the~~ above observations it is believed that items 8 and 21 may account for the difference between an effective and ineffective professor. Item 8 refers to the spirit of growth through presentation of fresh and vital materials as against the use of outlines of materials that haven't been brought up-to-date. Item 21 means that stimulus to thinking is created by the presentation of an abundance of novel and thought provoking viewpoints.

In view of our above observations we would restate our hypothesis to the effect that "aside from personality, scholarship, classroom administration, spirit of growth, and stimulus to thinking determine whether a professor is effective or ineffective."

TABLE II

## RANK OF THE CRITERIA ACCORDING TO DISCREPANCY SCORES

Rank	Discrepancy Score	Item Number
1.	6.60	1
2.	6.25	22
3.	5.90	8
4.	5.00	21
5.	4.75	17
6.	4.65	4
7.	4.65	9
8.	4.55	2
9.	4.50	25
10.	4.25	3
11.	3.90	15
12.	3.75	24
13.	3.60	10
14.	3.35	23
15.	3.00	11
16.	2.85	6
17.	2.65	18
18.	1.85	5
19.	1.45	14
20.	.65	13
21.	.55	7
22.	.30	12
23.	.25	16
24.	.20	20
25.	.15	

### Conclusions

In conclusion it should be pointed out that we don't know what "effective" and "ineffective" means with regard to professors. Neither do we know what personality factors are. However, personality factors could be operationally defined but would have little significance for the purpose at hand unless our definition corresponded with what persons mean when they use these concepts. In any case it would be important to consider whether personality is charismatic and something which arises from some innate quality or whether the "natural leader" type is inextricably associated with technical competence. We would hazard an observation that teachers are always in transition from the one to the other. Then, too, what people mean by effective and ineffective may well be based on what they want out of teaching.

### Suggestions for Further Research

1. Do further studies of the criteria of effective teaching in order to derive a more precise and conclusive picture of the effective teacher.
2. Determine these criteria on the basis of the most fundamental issues involved in teaching, i. e., in terms of the goals of education (more broadly), and in terms of the particular subject being taught (more specifically).
3. Do intensive and comprehensive (longitudinal) studies of professors in all fields of knowledge to see whether or not a sub-typology is necessary in order to evaluate the teaching of professors in different areas of teaching; whether social background factors of any specific organization can be seen to be more closely related to effective teaching than other life-history patterns; etc.
4. In studying the causal factors associated with effective teaching other factors must be looked into. Such factors as whether the teacher wanted to teach this particular subject, how long he has been teaching it, the composition of the class, etc.; these and many other factors should be considered.
5. Perhaps the rating of the effectiveness of the teaching of a particular teacher should be restricted to the effectiveness with which he teaches one particular course. Seminar courses must be differentiated from lecture courses, these from laboratory courses.
6. In using the rating of students as to the effectiveness of teachers one should consider how well that student is doing in the courses he is taking with that professor or has already taken and been graded one.
7. An approach suggested by current leadership studies may be to pursue the idea that an effective teacher is effective because of certain charismatic factors which one can't measure but somehow are known to exist. Or, as suggested by these studies, teaching effective-

May, 1954)

FRUMKIN - HOWELL

693

ness may be due to technical competence. Technical competence up to a certain point can be learned. That effective teaching can be learned is the view held by most experts in the search for the reasons for and the promotion of effective teaching.

# NOTES ON THE RATING OF TEACHER PERFORMANCE

DAVID G. RYANS  
University of California  
Los Angeles, California

## Teachers' Reactions to Ratings

TEACHERS DO not like to be judged; they do not look with favor upon proposals to assess their classroom performance. This seems to hold true for teachers of all grades or courses from elementary school to university. Organizations made up of teachers have gone on record as opposed to performance ratings in any form. The teaching profession probably is not unique. Most persons have an antipathy toward judgments of their behavior — unless they are assured that the results will be flattering.

Yet assessment or evaluation of performance is a sine qua non in any effective personnel program. Discriminating judgments of performance, and the subsequent assignment of ratings, are necessary wherever individual differences in performance are known to exist. Both the selection of personnel for effective job performance and the correction of deficiencies in job behavior of individuals already in service are dependent upon rating procedures.

Why are teachers, or perhaps most persons, so opposed to the idea of having their behavior judged and assessed?

There probably are several reasons. For example:

(1) There is distrust of those who serve as judges and assign the ratings. This is a commonly voiced criticism of behavior rating. The person being rated feels he will not be given a square deal. It is assumed that the rater is prejudiced and will employ the rating to reflect his bias.

(2) There is awareness of the much publicized unreliability of ratings. It is known that the ratings of untrained judges will show considerable disagreement. The ratings assigned by untrained persons may differ from one judge to another and for the same judge assigning ratings at different times. (It should be noted, of course, that in spite of the presumed unreliability of ratings, teachers regularly follow the practice, often with great assurance, of assessing various aspects of pupil behavior and assigning arbitrary ratings or "marks.")

(3) Teachers claim they are "professional" persons and therefore should not be subjected to such a prestige-weakening experience as performance evaluation. Professional status is presumed to preclude judgment of competency of job behavior.

(4) Many individuals, and many teachers in particular, are intro-

verted and mildly insecure. Therefore, they often are unduly fearful of the results of judgment of their behavior. They are afraid that judgments may be of an adverse nature and, therefore, consciously or unconsciously, they seek to avoid the potentially ego-damaging influence.

### Critique of Reactions to Ratings

There are, then, at least four reasons why teachers may not like to be rated. The last named, sensitivity, obviously does not constitute a valid criticism of performance ratings or a valid reason why they should not be used. Job performance cannot be improved, indeed it is likely to show progressive deterioration, if the issue of quality of behavior is avoided and standards of classroom performances are neglected.

Similarly, there is no justification for the view that enrollment in certain professional courses, leading to professional recognition in the form of a license or credential to practice the art of teaching, guarantees a high level of competency and precludes the need for judgment of quality of performance. Needless to say, the works of physicians, lawyers, ministers, and similar professional persons are judged by those they serve. If the lawyer does not win cases his practice suffers. The medical doctor who makes mistakes and "loses" too many patients finds himself without a practice. This is true of many other occupations as well as those just mentioned. The baseball player's record is very carefully kept and he is hired and fired on the basis of his performance. Similarly, the salesman's behavior is reflected in his sales record on the basis of which his success or lack of it is recognized. The building contractor or the merchant in normal times gains a reputation for a high or low quality product. Judgments are made in all walks of life and the practitioner recognizes that his performance must meet the standards that have been set.

So much for the last two reasons for teachers' unfavorable reactions to ratings. They may be dismissed without further comment.

The arguments (1) that judges or raters may be prejudiced and (2) that ratings sometimes are unreliable, may or may not be justified in a particular situation. But it is true that these criticisms describe conditions which could very well result in a distorted picture and could render judgments or ratings of limited usefulness for any important purpose.

### Two Methods of Objectifying Ratings

Two such procedures will be described briefly. One of these is the "Forced-Choice Performance Report,"<sup>1</sup> which has been used in con-

<sup>1</sup>Richardson, Marion W. "An Empirical Study of the Forced-Choice Perform-

nection with efficiency reports in the U. S. Army and also is coming into considerable use in industry. The other is a "Classroom Observation Scale,"<sup>2</sup> which has been used in research on the characteristics of teachers in connection with an American Council on Education project.

Essentially the forced-choice technique consists of presenting a pair of equally popular behavior descriptions (statements or phrases), one of which is known, on the basis of empirical validation, to discriminate between criterion groups, while the other does not. For example, two descriptions of teacher performance might be:

A. Gives "long distance" directions.

B. Tells pupils each step to take.

Both of these statements presumably describe an "autocratic" teacher and they might well be equally popular so far as frequency of their use in describing teacher behavior is concerned. Based upon pre-testing, however, one of these might discriminate successfully between "good" and "poor" teachers whereas the other did not. The judge, or rater, would not know which of the statements discriminated between the criterion groups. He would simply select the one that in his opinion most accurately described the teacher's behavior as he had observed it.

In practice, the forced-choice rating item is somewhat more complex. In the form used in the new Army efficiency report each item of behavior is covered by four statements or phrases, two of which are favorable and two unfavorable. In an industrial use of the forced-choice technique recently described by Richardson,<sup>3</sup> blocks of five statements were employed, each block consisting of: (a) a positively valid favorable statement; (b) a non-valid favorable statement, equally as popular as (a); (c) a negatively valid (usually unfavorable) statement; (d) a non-valid unfavorable statement equally as popular as (c); (e) a neutral, non-valid statement. The statements were arranged in random order and the rater was directed to select from each block the one statement most descriptive of the job performance of the individual and the statement least descriptive.

ance Report." (Paper presented at fifty-seventh meeting of the American Psychological Association, September, 1949.)

Sisson, E. Donald. "Forced-Choice -- The New Army Rating." Personnel Psychology, 1948, 365-381.

Staff, Personnel Research Section, Department of the Army. "The Forced Choice Technique and Rating Scales." The American Psychologist, 1946, 1, 267.

2. Staff, Teacher Characteristics Study. Classroom Observation Scale. (Confidential report form).

3. Richardson, Marion W. "An Empirical Study of the Forced-Choice Performance Report." (Paper presented at fifty-seventh meeting of the American Psychological Association, September, 1949.)

One of the items appearing in the efficiency report used in rating Army officers is shown below:

- A. Boastful
- B. Inspires pride in the organization
- C. Lacks tact
- D. Thoughtful of others

An item referring to teacher behavior might, for example, be found to follow some such form as the following:

- A. Gracious
- B. Approachable
- C. Fault-finding
- D. Cross

Neutral statements have been omitted in the above illustrations. The directions are simply to choose the description that is most applicable and the one that is least applicable to a particular individual.

The Classroom Observation Scale used by the Teacher Characteristics Study is basically an extension and refinement of instruments of this type that have been used previously. It has several unique features: (1) it provides for judgment of teacher behavior as based (a) on the immediate observation of the teacher's performance in the classroom and (b) on inferences regarding teacher behavior derived from pupil behavior; (2) it assumes that many teacher traits or qualities constitute dimensions of behavior, the opposite poles of which may be described with precise and meaningful terms referring to specific behaviors of the teacher; (3) it demands that the judge avoid the "central tendency error" by forcing the rating in the direction of one or the other of the poles; (4) it makes use of a detailed "Glossary" which provides supplementary descriptions of the teacher behaviors under consideration, and, of course, demands thorough acquaintance with the instrument.

Figures 1 and 2 show excerpts of the Classroom Observation Scale and accompanying Glossary.

The Classroom Observation Scale is one of the basic instruments used in the Teacher Characteristics Study. It was developed after an extensive and intensive critical review of all available investigations and discussions of teaching effectiveness, with particular attention to those reports which (1) sought to define the essential characteristics of teaching, (2) presented behaviors believed to be desirable in teachers and contributive to good teaching, (3) attempted to analyze basic personality structures emerging from human behavior.

Even more important in the development of the scale was a "critical incidents" study of teacher behaviors, leading to an extensive list of specific acts, or behaviors, believed by students, principals, supervisors, and training teachers to be indicative of effective or ineffective teaching.

**SCALE A: CLASSROOM OBSERVATION**  
**Teacher Characteristics Study**

Teacher \_\_\_\_\_ (Coded)  
 Date \_\_\_\_\_  
 Observer \_\_\_\_\_

School \_\_\_\_\_ (Coded)  
 Grade \_\_\_\_\_ Time \_\_\_\_\_  
 Subject \_\_\_\_\_

**PUPIL BEHAVIOR\***

1. Disinterested	1 2 3 4 0	Alert
2. Obstructive	1 2 3 4 0	Constructive
3. Restrained	1 2 3 4 0	Participating

**TEACHER BEHAVIOR\***

1. Partial	1 2 3 4 0	Fair
2. Autocratic	1 2 3 4 0	Democratic
3. Aloof	1 2 3 4 0	Responsive

Key: 1, 4 Markedly present

2, 3 Present

0 No opportunity for observation

\*See accompanying Glossary for definition

Figure 1. Representative Portions of Classroom Observation Blank.

**GLOSSARY**

To be used with Classroom Observation Study A  
Teacher Characteristics Study

Pupil BehaviorRestrained - Participatinga. Restrained:

- afraid to try
- halting
- unsure
- crying
- uncertain
- timid
- hesitant

b. Participating:

- willing to try
- undisturbed by mistakes
- volunteering
- entering freely into activity

Teacher BehaviorAutocratic—Democratica. Autocratic:

- giving "long distance" directions
- "laying down the law"
- telling pupils each step to take
- always the director rather than a participant

b. Democratic:

- entering into pupil activities without domination
- exchanging ideas
- encouraging pupil decision
- putting pupils "on their own"
- encouraging pupil initiative

Aloof—Responsivea. Aloof:

- cold
- apart
- removed
- condescending
- reserved

b. Responsive:

- warm
- commending effort
- encouraging
- gregarious
- tactful

Figure 2. Excerpts from Glossary of Classroom Observation Scale.

The Classroom Observation Scale underwent a number of revisions before the final form was adopted. Selection of the qualities for inclusion in the Scale was made after exhaustive review, as already noted, and in light of certain arbitrary criteria<sup>4</sup> agreed upon by the project staff.

The use of the Classroom Observation Scale is predicated on thorough training of observers who will employ the scale. In such training, value judgments must be deemphasized; the observer must learn to record what he sees rather than what he infers, and must think in terms of behavior as such instead of "good" or "poor" behavior.

#### Evaluation of the Suggested Methods

That carefully trained observers can reliably judge teacher behavior with the Classroom Observation Scale is indicated by obtained correlation coefficients of approximately .80 between the ratings of different judges, and correlation coefficients of like magnitude between the first and second ratings of the same observers.

In the use of either of the procedures suggested, Forced-Choice Ratings or the Classroom Observation Scale, it is, of course, necessary that the judges, or raters, (1) base their judgments on job performance, i.e., actual teacher behaviors. It is also necessary (2) that opportunity for observation be extensive enough to assume a representative sampling of the behavior being rated.

Naturally, the Classroom Observation Scale cannot eliminate the possibility of dishonest rating or error resulting from its unsophisticated use. But when used by impartial judges who have been trained in observing teacher behaviors and recording them on the Scale, it has been found to reflect teaching performance reliably, thereby giving lie to the argument that ratings should not be employed because of their essentially unreliable nature.

4. The criteria employed in selecting the qualities to be included in the Classroom Observation Scale were: (1) the trait or quality should be capable of identification in terms of observable teacher behavior, or observable pupil behavior, or known concomitants of either teacher or pupil behavior; (2) in general, the traits included should be mutually exclusive -- at least insofar as possible; (3) insofar as possible, traits included should be equally applicable to teachers in different school situations; (4) the traits included should be stated in terms for which the meanings are uniform to a high degree (i.e., there should be common understanding of the definitions of those trait names or terms employed); (5) the traits or qualities included should be ones that is generally agreed all good teachers should possess (or should not possess), or that all poor teachers probably do not possess (or probably do not possess).

In the forced-choice technique, the judge is forced to make ratings on the basis of observed teacher behavior rather than in light of a preconceived evaluation, since knowledge of a particular behavior description is lacking. Unless the scoring key is available to the rater (and, of course, it should not be) the possibility of dishonest and prejudiced rating is reduced to a minimum. Obviously, however, training in observation cannot be dispensed with by using the forced-choice approach. Training is always required if assessments are to possess reliability and validity.

### Organized Teacher Rating

On the assumption that teacher ratings are desirable and that they may be reliably and validly accomplished, one further suggestion is made that applies particularly to the rating of teachers — that, instead of dependent upon supervisors or principals for judgments of teacher behavior, highly selected and trained teams of judges be employed in a school system for that sole purpose. Under such a plan each teacher might be visited on several occasions throughout the year for the purpose of observation and assessment of teaching performance. The pooling of independent ratings of the judges would constitute a teacher's rating for a given period of time.

It is true, of course, that the responsibility of such judges in a school system would be tremendous and that their integrity would necessarily have to be above question. It is also true that the classroom observations would be artificial in a sense — that the teacher might attempt to put on a good show for the judges. Job performance is based upon rather thoroughly established habits, however, and it seems improbable that more than an occasional poor teacher would be a sufficiently good actor to fool adequately trained and experienced judges.

### Summary

- (1) Ratings of teacher performance are not only desirable but necessary if the school system fulfills its obligation of providing adequate instruction for its pupils.
- (2) The only plausible objections to ratings of teacher performance are (a) that raters may consciously or unconsciously reflect prejudice or bias and (b) that behavior rating is subjective and unreliable.
- (3) These "plausible" objections to teacher ratings will be met if the judgment of teacher behavior can be objectified.
- (4) The forced-choice rating technique is a promising device that may be useful in education as well as in the military service and in industry. It appears to satisfactorily objectify the rating process. With competent trained observers the Classroom Observation Scale,

May, 1954)

RYANS

70

as used in the Teacher Characteristics Study, also provides reliable objective and valid ratings of teacher behavior. The use of such instruments is recommended.

## BOOK REVIEWS

L. Joseph Lins, University of Wisconsin; Valworth Plum, University of Minnesota;  
John Schmid, Michigan State College

Burlington, Richard S.; May, Donald C. Handbook of Probability and Statistics with Tables (Sandusky, Ohio: Handbook Publishers, Inc., 1953), pp. ix+332.

Statisticians and research workers use a wide variety of mathematical tables and computational procedures in their work. The problem of finding whether and where appropriate tables are available is common to workers in many fields. The handbook cited above will be of value to workers in direct proportion to the extent to which their activities involve the use of mathematics beyond simple arithmetic. If their activities consist of collecting data with the view mainly of presenting them in tabular or graphical form, they will have infrequent occasion to refer to this source. However, if the research worker uses the mathematical theory of probability to determine with what confidence conclusions can be drawn and predictions can be made from the information in data collected in a specified way and analyzed in the appropriate manner, he will find the handbook very useful and likely come to regard it as indispensable.

The book contains two main divisions. Part I, pages 1-246, consists of the theoretical part, and working procedures, often with illustrative solutions of practical problems, covering much of the basic materials in probability and statistics. After a terse introductory chapter on the role of modern statistical methods including graphical representation, there is a chapter on fundamental definitions of elementary statistics. Then follow chapters on frequency distributions in one dimension, elementary probability theory, probability distributions in one dimension and generating and characteristic functions. The classical binomial, Poisson, and normal distributions have one chapter each.

Chapter XI is a highly informative presentation of probability distributions in two or more dimensions. Both discrete and continuous variables are treated, and probability theory is summarized for two and for three chance variables. A number of useful calculations with normal distributions are illustrated by graphical schemes. One such is a graphic method for testing normality by the plotting of observational data on normal probability graph paper. A chapter on regression theory and time series includes a discussion of fitting regression lines for the cases involving errors in ordinates, in abscissae, and in both ordinates and abscissae.

Chapter XIII is given to modern sampling distributions and their applications. Procedures in statistical inference including tests of significance and interval estimation are discussed and illustrated in Chapter XIV. There is a short but well-presented chapter on the analysis of variance for one and two-way classifications. Chapter XVI treats briefly but clearly various forms of interpolation theory, factorial polynomials, and finite differences. Chapter XVII contains clearly stated applications of statistical

theory in sequential analysis and quality control. The final chapter in Part I contains a short table of integrals and some fundamental mathematical functions and theorems.

Part II, pages 247-309, is comprised of twenty-three valuable tables. These tables are partly statistical and partly mathematical. The former include, tables on the classical distributions, such as the binomial and normal (several forms of each) and on the modern exact sampling distributions i.e., the F, Z, t, and  $\chi^2$  distributions. In addition there are tables of the incomplete beta function ratio and of the gamma function. It should also be pointed out that there are frequently short tables of various distributions given in some of the chapters on theory and application. One such short table, particularly useful in the social sciences gives for several values of  $\eta$ , the mean range,  $\bar{R}$ , and the standard deviation of  $R$  in units of  $(b - a)$  for a rectangular distribution. The mathematical tables include tables of factorials, logarithms of factorials, factorials and their reciprocals, squares, square-roots, and reciprocals, of natural trigonometric functions, and of natural and common logarithms of numbers.

About the only shortcomings the reviewer can suggest for such comprehensive work for the modern research worker are that it might have included additional statistical functions arising in multivariate analysis, power functions of various statistical tests, and certain nonparametric tests, with some of the more useful tables of these tests.

In conclusion, it can be stated that the authors have achieved to a marked degree their purpose in compiling a handbook that will serve well the needs of students and workers in statistics in many fields of application.

Palmer O. Johnson

University of Minnesota

Counts, George S., Education and American Civilization (New York: Bureau of Publications, Teachers College, Columbia University, 1952), pp. xv+491, \$3.75.

The central tenet of this volume is that education invariably expresses some conception of the civilization that it serves and consequently cannot rise above these concepts or transcend these limits. Counts points out that great education always expresses a great concept of civilization. Inasmuch as the central purpose of a democracy is the development of a great people, it follows that American education must constantly strive to reexamine its values to accurately express the American way of life as gleaned from America's past and present setting.

In order to insure a progressive development toward the optimum future, we must expend thought and energy in overcoming the lag between the movement of events and past loyalties and customs. Today, there is an urgent need for a great education. In the past, America thought that with education would come enlightenment and freedom, but the reverse has often proved the case. Tyranny has shackled education to serve despotism and attack the very foundation of our culture.

We in America have a rich heritage derived from many diverse cultures, a history of freedom and equality of opportunity, and a rich territorial setting. We have harnessed the forces of nature to do our bidding and have opened new realms of technology, yet we have failed to equip our minds for the demands of the new age.

American civilization is at a crossroads and it is our choice as to which values will survive in the new age. The best of our inheritance must be synthesized and embodied in our educational system to propagate our ideals. Among these American values are the Hebraic-Christian ethic, the humanistic spirit, the scientific method, the rule of law, and the democratic faith. The successful embodiment of these values in our framework is useful insurance to keep our civilization from falling into an abyss. This must be an education for a society of equal, free men, and must embrace the whole world community.

The author has developed his tenets with great care and consideration, and despite his misgivings noted in the preface, he has made this a work of significance in the literature of American education. His exhortations have much food for thought in ascertaining present trends, and defining goals. They are of significance in present curriculum revisions.

Counts has demonstrated a keen grasp of the significant influences on the American educational scene and has attempted a thoughtful extension of these basic doctrines into the future. However, the scene is changing so rapidly and the picture is so kaleidoscopic that it is difficult to ascertain if the progression outlined by the author would necessarily follow in the new civilization.

It can be said, nevertheless, that many of the paradoxes found in other works on the same subject have been successfully resolved in this work which presents a unified, though varied, account of the movement of forces affecting the educational scene.

The social scientist may perhaps disagree with the points of emphasis and the educator may find fault with stress on the mundane, but the average reader's knowledge may well be strengthened and his thinking stimulated through the reading of this volume.

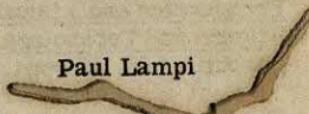
The test is written in a genre that all may grasp. Ideas are carefully developed throughout.

Duluth Public Schools  
Duluth, Minnesota

Kenworthy, Leonard S. World Horizons for Teachers (New York: Teachers College, Columbia University, 1952), pp. 160, \$3.25.

At a time when self-anointed super-patriots are trying to take the E out of UNESCO, it is encouraging to read a book designed to help instill world-mindedness in teachers. Kenworthy states the purpose of his book

Paul Lampi



in unequivocal terms: "The times demand an educational crusade for world-community, and the teacher needs to be in the forefront of this crusade."

The author outlines specific ways in which teachers may cultivate world-mindedness. The teacher must be (1) an integrated individual; (2) an expert in democratic human relations; (3) rooted in his own country and culture; (4) appreciative of other countries and cultures; (5) an informed participant in efforts to strengthen the United Nations and to achieve world community; (6) conversant with methods and materials for creating world-minded children and youth; and (7) buttressed by a faith or a philosophy of life to undergird his mission of developing world-mindedness in himself and others.

No book can say everything. But perhaps the criticism should be made that Kenworthy glosses over the maze of difficult problems which face our new atomic age in a period of Soviet-American tension. By passing lightly over the many Herculean difficulties that vex our international leaders, the book may tend to develop tender-minded naivete on the part of students. Rather it should develop citizens who are fully aware that present world horizons are clouded with numerous perplexing enigmas. This criticism may result from the fact that Kenworthy approaches the problem of world-mindedness with missionary zeal rather than with scientific objectivity.

The book contains a good bibliography both of general works and of specific teaching materials. Unfortunately these bring us up to only about 1949. For example, the bibliography lists F. C. S. Northrop's The Meeting of East and West (Macmillan, 1946), but says nothing of Northrop's two later and equally pertinent books, Ideological Differences and World Order (Yale University Press, 1949) and The Taming of the Nations (Macmillan, 1952).

University of Minnesota, Duluth Branch

Henry Ehlers

West, Michael. A General Service List of English Words with Semantic Frequencies and a Supplementary Word-List for the Writing of Popular Science and Technology. (London: Longmans, Green and Co., 1953), pp. xiii + 588, 50/-net.

The General Service List of English Words is the culmination of a long effort to find a relatively short list of words to be used as a basis for teaching and learning English as a foreign language. In 1934 the Carnegie Corporation under the leadership of F. P. Keppel sponsored a conference in New York City to discuss the problem of word-lists. The chairman elected was Robert Herndon Fife, for many years Chairman of the Committee on Modern Languages of the American Council on Education. It seemed to this conference that the time had come for coordinating the efforts of individual workers, and a committee consisting of L. W. Faucett, H. E. Palmer, and Michael West, with E. L. Thorndike as consultant, was appointed to prepare a report. The result was the issuance in 1936 of a

tentative General Service List of approximately 2000 words as Part V of the Interim Report on Vocabulary Selection. "Education Authorities throughout the world. . . in active contact with the problems of English teaching" were "cordially invited to experiment with it, to criticize it, and to suggest any changes or development which would enhance its value." The 1953 list is a revision by West of this early list.

The General Service List grew out of an attempt to simplify the learning of English rather than to develop a simplified form of English for universal use. As G. B. Jeffrey points out in the Foreword to the revised List, "A language is so complex that selection from it is always one of the first and most difficult problems of anyone who wishes to teach it systematically. . . To find the minimum number of words that could operate together in constructions capable of entering into the greatest variety of contexts has been therefore the chief aim of those trying to simplify English for the learner."

West has made use of the Lorge-Thorndike Semantic Count of English Words to suggest the relative importance of various meanings and uses of the words listed. When in the editor's judgment the meanings are so different that they are equivalent to separate words for purposes of teaching, that fact is indicated. Indicated also are items which should not be taught or on which the editor is doubtful. Meanings are given in terms of definitions or illustrative expressions or both. Differences between American and British usage are pointed out.

The following is the entry for the word immediate:

IMMEDIATE, 290e	?	(1) ( <u>without intermediary, direct</u> ) My im-	
adj.		mediate aim is. . . Immediate evidence,	
		knowledge - not hearsay or reasoning	21%
	?	(2) ( <u>very near</u> ) Immediate companion,	
		neighbors. The immediate future	26%
	(3)	( <u>Without delay</u> ) An immediate reply.	
		For immediate use	52%

Thus the listing shows that immediate is an adjective and that it occurred an estimated 290 times in a count of five million words (actually 145 times in two-and-a-half million words). In 21 per cent of these occurrences it meant without intermediary, direct. The editor suggests however that this use of the word may be omitted by the teacher.

As West points out, the list is based on printed and written material and for that reason tends to give less importance than is deserved to some items of speech. The point is made also that frequency of use is only one criterion for word selection. For these reasons the judgment of such a scholar as West has special value.

The General Service List of English Words is an important contribution to the teaching of English as a foreign language. It will be useful whether British or American English is to be taught. The reviewer recommends it for students of language teaching, for textbook writers, and for teachers. At the high school and college level it will be helpful to students themselves

after they have gone beyond the first stages of their learning. The list has value also for the teaching of English as a native language. It provides a simplified form of the Lorge-Thorndike Semantic Count on a large number of useful words. The General Service List should take its place with the lists of Thorndike, Horn, and others, as a significant resource in language teaching.

Herschel T. Manuel

University of Texas

Traxler, Arthur E.; Jacobs, Robert; Selover, Margaret; Townsend, Agatha.  
Introduction to Testing and the Use of Test Results in Public Schools  
(New York: Harper and Bros., 1953), pp. 113, \$2.50.

This book was designed to provide basic information relative to the use of tests in understanding children. The material covers the subject comprehensively, yet in concise manner deals with various pertinent and practical problems confronted by school people from the time group evaluation begins until final test results are used to assist individual pupils.

The selection of chapter topics impresses one with the fact that this book was prepared by people who understand the major problems teachers and administrators meet in using tests. The authors build a solid case for the point of view that tests should augment other information available in any school, stressing the need for the school to secure at least five kinds of information on all pupils: general aptitudes or abilities, specific aptitudes or abilities, achievement in different fields of study, educational and vocational interests (and here could be added, avocational interests), and personal and social adjustment data. Tests may assist in securing such information.

Some attention to the specific limitations of tests is presented in timely and emphatic fashion as a guard against the enthusiasm, perhaps too frequently found in some school situations, where test results alone are used out of proper perspective in making generalizations about pupils.

The philosophy that teachers should have an active part in planning the testing program, in selecting tests, and in administering and using test results is a practical and sensible point of view. This discussion gives in "suggerated" form information which teachers need in order to function in the area of psychometrics. This information, when presented in such understandable fashion, meets a great need in public schools as, in the past, many teachers have lacked the theoretical and practical understandings essential to use tests in optimum ways.

A particular strength of this volume is the final chapter in which a particular case is used to illustrate how the results of many tests, along with other data the school accumulates, may assist teachers to help a given boy make good use of his school experience.

The general format of the book ( $8\frac{1}{2} \times 11$  in size) makes it easily read. It is well illustrated with graphs and tables in portions which require pic-

ture material to clarify the printed page.

There are lists of suggestions for further related reading following each of the ten chapters. An appendix furnishes a list of major test publishers.

Although the volume will probably find its greatest use as a basic text in teacher training where consideration is given to educational measurement and pupil guidance, it should meet a recognized need of many school counselors and administrators for a practical reference.

Carl H. Waller

Director, Child Study and Service  
Madison Public Schools  
Madison, Wisconsin

#### Brief Commentaries

Althouse, J. G. Structure and Aims of Canadian Education (Toronto: W. J. Gage and Co., 1949), pp. 77, \$1.25.

This book presents two lectures of Dr. Althouse, Chief Director of Education for Ontario. The first lecture treats the administrative structure of Canadian education as an outgrowth of a provincial heritage and responsibility. The second lecture discusses the dominant philosophy of Canadian secondary education as an institution for selecting and training the brightest brains versus a practical education for all, discipline, and the development of the students' social characteristics. - J. S.

Baker, Harry J. An Introduction to Exceptional Children (New York: Macmillan Co., 1953), pp. 500, \$5.00.

The first edition to this book was published nearly a decade ago. The author feels that the change from a wartime era to a peacetime era has resulted in a redirection of emphasis in education. This book is intended for use in an introductory course for college and university students and for teachers who are attempting to specialize in some area of exceptional children. The following topics are taken up: the physically handicapped; mental growth and development; neurological and psychogenic diseases; behavior adjustments; and educational retardation. - J. S.

Moustakas, Clark E. Children in Play Therapy (New York: McGraw-Hill, 1953), pp. vii+218, \$3.50.

This book furnishes to psychologists, teachers, and parents a variety of situational experiences in play-therapy with both well-adjusted and dis-

turbed children. It attempts to show how specialized play situations may be used in a therapeutic manner to help children attain emotional stability and maturity. - J. S.

Percival, W. P. Should We All Think Alike (Toronto: W. J. Gage Co., 1951), pp. 112, \$1.25.

Should We All Think Alike is a continuation of a lecture series, the first of which was "The Structure and Aims of Canadian Education" by J. G. Althouse. The primary topic discussed in this book is the differentiating characteristics of French-Canadian education in Quebec. Its discussion is subsumed under three topics: the problem of multiplicity of thought; education at the doorstep of youth; and a nation, harmonious and strong. - J. S.

Robbins, Florence Greenhoe. Educational Sociology (New York: Henry Holt and Co., 1953), pp. 529, \$4.75.

This book presents a comprehensive approach to the topics of the social and cultural orientation of the child; the social and cultural aspects of the school; and the integration of the child, school, and community. - J. S.

Spears, Harold. Improving the Supervision of Instruction (New York: Prentice-Hall, Inc., 1953), pp. xii+478, \$4.75.

As background for his presentation of the improvement of supervision, Spears reports at length what has been and what is practice in supervision in elementary and secondary schools of the United States. In this he acquainted himself with the Departments of Education of the 48 states and surveyed supervisory programs of 150 city and county school systems. Especially impressive in this book are the hints on how to improve the supervisor-teacher relationship and on methods for improving the teacher, supervisor, and instruction. - L. J. L.

Tyler, Leona F. The Work of the Counselor (New York: Appleton-Century-Crofts, Inc., 1953), pp. x+323, \$3.00.

Having in mind primarily the use of this book for a beginning course for psychology students and for persons expecting to be doing some guidance work in high schools or colleges, Tyler has used a generalized approach to counseling rather than pointing the discussion to any specialized group of counselors. - L. J. L.

## Research News and Communications

Lawrence P. Blum, University of Wisconsin Extension, Milwaukee; Jacob O. Bach,  
Southern Illinois University; Lester Sands, Santa Barbara College

Report on Annual Meeting of California Council on Educational Research.  
The annual meeting of the California Council on Educational Research was held on November 6 and 7 in Santa Barbara, California. In attendance were 140 superintendents, principals, teachers, directors of research, and college and university professors. The group is led by a committee of nine directors and sponsored jointly by the California Teacher's Association and the State Department of Education.

Considerable research was summarized through a series of panels. The four major topics covered were: Education of the Gifted, Improvement of Test Instruments, Evaluation of Special Areas of Education and Methodology of Recent Research.

The group dealing with the topic "Education of the Gifted," brought to light the various ways that many communities have found successful in educating those who have superior endowments. Considerable attention is being directed to these children in California at the present time. The dire need for additional research in this area brought to light the fact that the council directors have submitted a request for approximately one and one half million dollars from a well known foundation for funds to underwrite some of the needed research in this area.

The panel dealing with the topic "Improvement of Test Instruments," covered two major topics - local and regional test norms and the non-test instruments of evaluation.

Speaking at the two luncheon meetings were G. T. Buswell of the University of California and W. H. Cowley of Stanford University. Dr. Buswell emphasized three ways of improving educational research. "The design of research studies needs more careful planning," he said. "There is a difference between tinkering and experimenting; between a well-planned investigation and a collection of facts."

His second suggestion for improvement is the use of "better techniques and instruments than are often employed. Too many of our studies have not gone beyond the statistics of means and medians."

Finally, "our research should be directed in greater volume toward crucial problems of education and should reflect current thinking about education."

"Just as nuclear research has a dominant center of interest in physics and chemistry, so research in education should be aware of what are the real problems of each generation. Too many studies are still dealing with issues which were real in 1920 but which are no longer of great concern," he said.

Buswell pointed out he recognizes that a "certain specialization and narrow definition" is necessary in searching for truth. But "unless those doing research keep abreast of the issues in education, their contributions will have little value for improving education." Admitting this puts a "double burden on the researcher," he said, "I see no escape from it."

"Unless research keeps abreast with important issues in education," he said, "the practices in our schools will be based on opinion rather than be guided by the findings of competent studies."

Dr. Cowley emphasized there two kinds of educational research. One is dealing with important conceptual problems; the other deals with facts.

"I am convinced the conceptual research is more important," the professor declared. "It causes me worry that many of these larger issues are not being touched by educational research." At the same times, "the conceptual type of research is not in conflict with the deep and important tradition of factual research. The two are supplementary."

"The accumulation of facts" — through factual research — "is nothing unless it is related to an important concept, or theory, or 'leading idea,'" he said.

Cowley also touched upon what he termed "academic freedom rows" which have been frequent in recent years. He pointed out, that there have been precisely the same "academic freedom rows" for centuries, and "when we stop having them we should be worried." But here, he said, is another example of a problem to be tackled with "idea research" or conceptual research.

"We can't do much about this factually without some leading ideas," he stated.

The next meeting of the research council is scheduled to be held in northern California, March 1, 1954. Additional information can be secured from G. W. Durflinger, Santa Barbara College, Santa Barbara, California, who submitted the report.

Study Effects of Counseling Probationary Students. The study at the University of Iowa was undertaken in an attempt to ascertain, under controlled conditions, the relationship of academic advisement to the scholastic performance of failing college students. The results suggested that: one hour seems to be as effective as four hours of academic advisement in producing improvements in the scholastic achievement of selected failing Liberal Arts College students at Iowa; a systematic advisory program seems to be associated with a level of scholarship significantly higher than that for comparable uncounseled students, under the conditions of this study.

These findings indicate the desirability of a formal guidance program for probationary students, featuring one relatively short, highly structured interview with a person identified with the office administering probationary procedures. Presumably, such a program would be effective for individuals selected in the same manner as the subjects in this experiment and in two preliminary studies, and who are in residence at institutions employing admissions practices similar to those in force at the State University of Iowa.

The study was made by E. L. Klingelhofer as a Ph. D. dissertation sponsored by D. B. Stuit and H. P. Bechtoldt.

Cooperative Graduate Summer Sessions in Statistics. Beginning in 1954 North Carolina State College, the University of Florida, Virginia Polytechnic Institute, and the Southern Regional Education Board will jointly sponsor cooperative Graduate Summer Sessions in Statistics.

The first session will be conducted by a distinguished faculty at Virginia Polytechnic Institute in the summer of 1954. Additional summer sessions are tentatively planned for North Carolina State College and the University of Florida in the two following years. Subsequent sessions will be rotated among these or other institutions throughout the south.

The summer sessions are designed to carry out a recommendation of the Southern Regional Education Board's Commission on Statistics, on which the three institutions initiating the program are represented. They will be of particular interest to (1) research and professional workers who want intensive instruction in basic statistical concepts and who wish to learn modern statistical methodology; (2) teachers of elementary statistical courses who want some formal training in modern statistics; (3) prospective candidates for graduate degrees in statistics; (4) graduate students in other fields who desire supporting work in statistics; and (5) professional statisticians who wish to keep informed of advanced specialized theory and methods.

Inquiries should be addressed to Boyd Harshbarger, Head, Department of Statistics, Virginia Polytechnic Institute, Blacksburg, Virginia.

## Research Abstracts and Bibliographies

T. A. Lamke, Iowa State Teachers College, Cedar Falls, Iowa

Horkheimer, Mary F. and John W. Diffor, editors. Educators Guide to Free Films. (Randolph, Wis.: Educators Progress Service, 1953), 516pp., \$6.00.

Lists 2574 titles of free motion picture films. Gives title, description of contents, size, whether sound or silent, number of reels, date of release, running time, names and addresses of distributors and their branch offices, and limitations of distribution. Contains title index, subject index, and source and availability index. Includes a paper by Dean John Guy Fowlkes on "The Significance of Films in Curriculum Improvement."

Horkheimer, Mary F. and John W. Diffor, editors. Educators Guide to Free Slidefilms. (Randolph, Wis.: Educators Progress Service, 1953), 185 pp., \$4.00.

Lists 621 titles of free filmstrips. Gives title, description of contents, number of frames included, source, and other pertinent information. Contains title index, source index, subject index.

Lazar, May, editor. The Retarded Reader in the Junior High School. Bureau of Educational Research Publication No. 21. (New York: City Board of Education, 1952), 126 pp., n.p.

Discusses the special reading problem in the junior high school. Suggests provisions for a special reading program and includes materials which may be used in such a program. Contains extensive bibliographies including selected references for teachers, selected book lists helpful to teachers of reading and materials for pupils.

Muller, Sister Mary Janet. General Education in the American Catholic Secondary School. (Washington, D.C.: Catholic University of America Press, 1952), 104 pp., \$2.00.

Discusses origins and the meaning for the present of general education. Contains a 150 item bibliography.

Personnel-O-Gram. Vol. 7, No. 4. (Lexington, Ky.: American College Personnel Association, 1953), 82 pp., n.p.

Contains proceedings of the 1953 annual convention. Contains summaries of numerous papers dealing with college personnel problems.

Polley, John W. and others. Community Action for Education. Institute of Administrative Research Study No. 9. (New York: Teachers College Columbia University, 1953), 102 pp., \$1.75.

"Reports ways and means which have proved successful in developing a sense of community in an area of a great city, in discovering lay leaders, and in making a significant beginning at strengthening creative interrelationships between great city schools and their immediate publics." Deals with the Bronx Park area in New York City.

Progress Report: Research on the Psychology of Reading. (Philadelphia, Penna.: Reading Clinic, Department of Psychology, Temple University, 1953), 4 pp., n.p.

Lists studies made in cooperation with the clinic. Contains 40 studies which have been completed or are in progress.

Ricciuti, Henry N. and John W. French. Development of Personality Tests for Naval Officer Selection. II. Validation of Experimental Tests at U. S. Naval Academy. Technical Report No. 2. (Princeton, N.J.: Educational Testing Service, 1952), 39 pp., n.p.

Describes efforts to develop and validate personality tests useful in the selection of naval officers. Results were not of immediate practical value but show promise.

Roe, Anne. A Psychological Study of Eminent Psychologists and Anthropologists and a Comparison with Biological and Physical Scientists. Psychological Monographs No. 352, Vol. 67, No. 2. (Washington, D.C.: American Psychological Association, 1953), 55 pp., \$1.50.

The third and last monograph in a series of clinical studies of research scientists. The series was designed to investigate the existence of relationships between life history, intellectual functions or personality characteristics, and the selection and pursuit of a particular science as a profession.

Shosteck, Robert. Five Thousand Women College Graduates Report. (Washington, D.C.: B'nai B'rith Vocational Service Bureau, 1953), 66 pp., n.p.

A study based on a nation-wide sample of women graduates of liberal arts colleges of 1946-49. Discusses pre-college background, reasons for college attendance, college curricula selected, vocational plans, and post-college activities.

Statistical Summary of Education, 1949-50. Biennial Survey of Education in the United States, 1948-50, Chapter I. (Washington, D.C.: Government Printing Office, 1953), 52 pp., 20¢.

Published separately, this first chapter of the latest biennial survey provides an abstract together with brief interpretative comments, of educational statistics dealing with number, size, and financial support of public schools, composition and size of enrollments, and kind and frequency of curricular offerings.

Wessel, Rosa and Goldie B. Faith. Professional Education Based in Practice. (Philadelphia, Penna.: University of Pennsylvania School of Social Work, 1953), 170 pp., n.p.

Two studies dealing with the place of practice in education for social work and the relationship of the professional school to the social agency. Contains a 60-item bibliography.

Wheeler, E. G. and D. F. Showalter. Better Teaching and Better Learning in the Social Studies. (Manhattan, Kans.: Kansas State College Press, 1953), 55 pp., n.p.

The report of a program of action research at the high school level.

Wilcox, Nancy J. Educational Travel Courses for Teachers. (Dansville, N.Y.: F.A. Owen Publishing Co., 1953), 40 pp., n.p.

A comprehensive study of the development and present status of educational travel courses in colleges and universities in the United States. Contains suggestions for planning and organizing travel courses.

Baird, Norman B. Educational Finance and Administration for Ontario. Bulletin No. 14 (Toronto, Canada: University of Toronto, Ontario College of Education, Department of Educational Research, 1952). 33 pp., 50¢

A survey of the practices used in the Canadian provinces, and in other countries, in the distribution of state aid to education. Contains a selected bibliography of 18 items.

Barrington, Thomas M. The Introduction of Selected Educational Practices into Teachers Colleges and Their Laboratory Schools. (New York: Bureau of Publications, Teachers College, Columbia University, 1953). 112 pp., \$2.50.

A study of the character and rate of diffusion of new educational practices among publicly-supported teachers colleges in the United States. The findings are not unlike those by Mort and Cocking for the public schools. Contains a 70-item bibliography.

Developing Children's Power of Self-Expression through Writing. Curriculum Bulletin 1952-53 Series, No. 2. (New York: Board of Education of the City of New York, 1953). 171 pp., n.p.

An outline of field research followed by a year of work in setting down the techniques tried and the results obtained, this bulletin is concerned largely with teaching children to write in a way that will stimulate personal reactions, develop the power of self-expression, and encourage critical thinking. Contains 60 selected references.

Education Abstracts is a periodical published ten times a year by the Education Clearing House of UNESCO. Annual subscriptions are \$1.75, and are available in the U.S. from the Columbia University Press, 2960 Broadway, New York, 27. Material is abstracted which is of particular interest to field workers, research students, and others engaged or interested in

fundamental and adult education.

The Eyes Get It. (Plainsfield, N. J., Interstate Press, 1953). 60 pp., n. p.

The 1953 yearbook of the New Jersey Secondary School Teachers Association deals with some of the most outstanding practices in the field of visual education that are now being employed by individual classroom teachers throughout the State of New Jersey. Contains a 29-item bibliography.

Gray, William S. Preliminary Survey on Methods of Teaching Reading and Writing. Educational Clearing House, UNESCO, Educational Studies and Documents, July, 1953, No. V, Parts I and II. (New York: Columbia University Press, 1953). 66 and 72 pp., gratis.

A survey analyzing and describing the various methods used in teaching both children and adults to read and write their native tongue. Considers implications for the improvement of the teaching of reading and writing, with major emphasis on the adult level.

Elementary Teachers Guide to Free Curriculum Materials. Tenth Annual Edition, 1953. (Randolph, Wisconsin: Educators Progress Service, 1953) 334 pp., \$4.50.

Lists 1242 items available from manufacturers, chambers of commerce, state and federal agencies, and the like. Provides illustrative teaching units making use of the sample materials. Contains indexes by title, source, and subject.

Jersild, Arthur T., and others. Education for Self-Understanding. (New York: Bureau of Publications, Teachers College, Columbia University, 1953). 54 pp., 85¢.

Report of a high school teachers' conference dealing with the teaching of psychological principles at the high school level. Contains selected references.

Kinney, Mary R. Bibliographical Style Manuals: A Guide to Their Use in Documentation and Research. A.C.R.L. Monographs, No. 8. (Chicago: Association of College and Reference Libraries, 1953), 21 pp., 60¢.

A survey of the most pertinent sources of information on bibliographical citation in the humanities, social sciences, and sciences. Concise and meaty, this monograph should prove invaluable to the teachers of courses in research, to the editors of scholarly publications, and to reference librarians.

Lambertson, Eleanor C. Nursing Team Organization and Functioning. (New York: Bureau of Publications, Teachers, Teachers College, Columbia University, 1953). 89 pp., \$2.25.

Presents a method of utilizing nursing personnel in hospitals most efficiently. Possibly something similar could be done for classroom teachers. Contains a 64-item bibliography.